

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
13 June 2002 (13.06.2002)

PCT

(10) International Publication Number
WO 02/45495 A2

- (51) International Patent Classification⁷: A01K 67/027
- (21) International Application Number: PCT/US01/46864
- (22) International Filing Date: 5 December 2001 (05.12.2001)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
- | | | |
|------------|-------------------------------|----|
| 60/251,819 | 6 December 2000 (06.12.2000) | US |
| 60/254,314 | 8 December 2000 (08.12.2000) | US |
| 60/255,968 | 13 December 2000 (13.12.2000) | US |
| 60/255,972 | 13 December 2000 (13.12.2000) | US |
| 60/255,983 | 13 December 2000 (13.12.2000) | US |
| 60/256,232 | 13 December 2000 (13.12.2000) | US |
- (63) Related by continuation (CON) or continuation-in-part (CIP) to earlier applications:
- | | |
|----------|-------------------------------|
| US | 60/251,819 (CON) |
| Filed on | 6 December 2000 (06.12.2000) |
| US | 60/254,314 (CON) |
| Filed on | 8 December 2000 (08.12.2000) |
| US | 60/255,968 (CON) |
| Filed on | 13 December 2000 (13.12.2000) |
| US | 60/255,972 (CON) |
| Filed on | 13 December 2000 (13.12.2000) |
| US | 60/255,983 (CON) |
| Filed on | 13 December 2000 (13.12.2000) |
| US | 60/256,232 (CON) |
| Filed on | 13 December 2000 (13.12.2000) |
- (71) Applicant (for all designated States except US): DELTA-GEN, INC. [US/US]; 740 Bay Road, Redwood City, CA 94063 (US).
- (72) Inventors; and
- (73) Inventors/Applicants (for US only): ALLEN, Keith, D. [US/US]; 224 Custer Trail, Cary, NC 27513 (US). LEVITEN, Michael, W. [US/US]; 3166 Bryant Street, Palo Alto, CA 94306 (US). KLEIN, Robert [US/US]; 1044 Webster Street, Palo Alto, CA 94301 (US).
- (74) Agents: BURKE, John, E. et al.; Deltagen, Inc., 740 Bay Road, Redwood City, CA 94063 (US).
- (81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.
- (84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).
- Published:
— without international search report and to be republished upon receipt of that report
- For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.



WO 02/45495 A2

(54) Title: TRANSGENIC MICE CONTAINING TARGETED GENE DISRUPTIONS

(57) Abstract: The present invention relates to transgenic animals, as well as compositions and methods relating to the characterization of gene function. Specifically, the present invention provides transgenic mice comprising mutations in a CHD2-like gene, a TDCDF1 gene, a TR2 gene, an E3-like gene, a CSPI gene, and an Abcd2 gene. Such transgenic mice are useful as models for disease and for identifying agents that modulate gene expression and gene function, and as potential treatments for various disease states and disease condition.

TRANSGENIC MICE CONTAINING TARGETED GENE DISRUPTIONS**Field of the Invention**

5 The present invention relates to transgenic animals, compositions and methods relating to the characterization of gene function.

Background of the Invention

10 A key to finding treatments for many human disorders has been the development of appropriate animal models to aid in the study of gene function. Functional analysis of genes, or functional genomics, frequently involves the introduction of modified genes into the germline, thereby generating transgenic animals. Gene targeting has also been used in various systems, from yeast to mice, to introduce site-specific mutations in the genome. Examining the effects of mutations within specific genes has proven to be one of the most productive and insightful approaches to
15 functional genomic analysis. This approach involves correlating mutations within specific genes to phenotypes or disease conditions that result from those mutations.

 With increasing awareness that mouse mutations can provide useful insights about the function of human genes, a great deal of interest has developed in systematically generating mutations within mouse genes that correspond to important human genes. The generation of knockout mice
20 using homologous recombination in embryonic stem cells is a powerful tool for such investigations. The understanding of many important mechanisms, including those involved in cell division, differentiation, aging and death, has dramatically advanced through exploitation of knockout technology.

 The targets of functional genomic approaches include those implicated in normal
25 development and disease, such as growth factors, growth factor receptors, signal transduction molecules, enzymes, cytoplasmic proteins, secreted proteins and nuclear proteins. Such genes a proven history as therapeutic targets. For example, over the past 15 years, nearly 350 therapeutic agents targeting G-protein coupled receptors involved in signal transduction have been successfully introduced onto the market. A clear need exists for identification and characterization of genes that
30 can play a role in preventing, ameliorating or correcting dysfunctions or diseases.

Summary of the Invention

 The present invention generally relates to transgenic animals, as well as to compositions and methods relating to the characterization of gene function.

 The present invention provides transgenic cells comprising a disruption in a targeted gene.
35 The transgenic cells of the present invention are comprised of any cells capable of undergoing

homologous recombination. Preferably, the cells of the present invention are stem cells and more preferably, embryonic stem (ES) cells, and most preferably, murine ES cells. According to one embodiment, the transgenic cells are produced by introducing a targeting construct into a stem cell to produce a homologous recombinant, resulting in a mutation of the targeted gene. In another
5 embodiment, the transgenic cells are derived from the transgenic animals described below. The cells derived from the transgenic animals includes cells that are isolated or present in a tissue or organ, and any cell lines or any progeny thereof.

The present invention also provides a targeting construct and methods of producing the targeting construct that when introduced into stem cells produces a homologous recombinant. In one
10 embodiment, the targeting construct of the present invention comprises first and second polynucleotide sequences that are homologous to the targeted gene. The targeting construct also comprises a polynucleotide sequence that encodes a selectable marker that is preferably positioned between the two different homologous polynucleotide sequences in the construct. The targeting construct may also comprise other regulatory elements that may enhance homologous recombination.

15 The present invention further provides non-human transgenic animals and methods of producing such non-human transgenic animals comprising a disruption in a targeted gene. The transgenic animals of the present invention include transgenic animals that are heterozygous and homozygous for a mutation in the targeted gene. In one aspect, the transgenic animals of the present invention are defective in the function of a down syndrome cell adhesion molecule-like (CHD2-like)
20 gene, a thymic dendritic cell-derived factor 1 (TDCDF1) gene, a toll-like receptor 2 (TR2) gene, a ubiquitin ligase E3-like (E3-like) gene, a contrapsin-like serine protease inhibitor (CSPI) gene, or an ABC transporter ATPase 2 (Abcd2) gene. In another aspect, the transgenic animals of the present invention comprise a phenotype associated with having a mutation in a CHD2-like gene, a TDCDF1 gene, a TR2 gene, an E3-like gene, a CSPI gene, or an Abcd2 gene.

25 The present invention also provides methods of identifying agents capable of affecting a phenotype of a transgenic animal. For example, a putative agent is administered to the transgenic animal and a response of the transgenic animal to the putative agent is measured and compared to the response of a "normal" or wild type mouse, or alternatively compared to a transgenic animal control (without agent administration). The invention further provides agents identified according to such
30 methods. The present invention also provides methods of identifying agents useful as therapeutic agents for treating conditions associated with a disruption of the targeted gene.

The present invention further provides a method of identifying agents having an effect on expression or function of a target gene. The method includes administering an effective amount of the agent to a transgenic animal, preferably a mouse. The method includes measuring a response of the
35 transgenic animal, for example, to the agent, and comparing the response of the transgenic animal to a control animal, which may be, for example, a wild-type animal or alternatively, a transgenic animal

control. Compounds that may have an effect on target gene expression or function may also be screened against cells in cell-based assays, for example, to identify such compounds.

The invention also provides cell lines comprising nucleic acid sequences of a target gene. Such cell lines may be capable of expressing such sequences by virtue of operable linkage to a promoter functional in the cell line. Preferably, expression of the target gene sequence is under the control of an inducible promoter. Also provided are methods of identifying agents that interact with the targeted gene, comprising the steps of contacting the target gene with an agent and detecting an agent/target gene complex. Such complexes can be detected by, for example, measuring expression of an operably linked detectable marker.

The invention further provides methods of treating diseases or conditions associated with a disruption in a targeted gene, and more particularly, to a disruption in the expression or function of the targeted gene. In a preferred embodiment, methods of the present invention involve treating diseases or conditions associated with a disruption in the targeted gene's expression or function, including administering to a subject in need, a therapeutic agent that effects target gene expression or function. In accordance with this embodiment, the method comprises administration of a therapeutically effective amount of a natural, synthetic, semi-synthetic, or recombinant CHD2-like genes, TDCDF1 genes, TR2 genes, E3-like genes, CSPI genes, or Abcd2 genes; CHD2-like gene products, TDCDF1 gene products, TR2 gene products, E3-like gene products, CSPI gene products, or Abcd2 gene products, or fragments thereof as well as natural, synthetic, semi-synthetic or recombinant analogs.

The present invention further provides methods of treating diseases or conditions associated with disrupted targeted gene expression or function, wherein the methods comprise detecting and replacing through gene therapy mutated target genes.

Definitions

The term "gene" refers to (a) a gene containing at least one of the DNA sequences disclosed herein; (b) any DNA sequence that encodes the amino acid sequence encoded by the DNA sequences disclosed herein and/or; (c) any DNA sequence that hybridizes to the complement of the coding sequences disclosed herein. Preferably, the term includes coding as well as noncoding regions, and preferably includes all sequences necessary for normal gene expression including promoters, enhancers and other regulatory sequences.

The terms "polynucleotide" and "nucleic acid molecule" are used interchangeably to refer to polymeric forms of nucleotides of any length. The polynucleotides may contain deoxyribonucleotides, ribonucleotides and/or their analogs. Nucleotides may have any three-dimensional structure, and may perform any function, known or unknown. The term "polynucleotide" includes single-, double-stranded and triple helical molecules.

"Oligonucleotide" refers to polynucleotides of between 5 and about 100 nucleotides of single- or double-stranded DNA. Oligonucleotides are also known as oligomers or oligos and may be

isolated from genes, or chemically synthesized by methods known in the art. A "primer" refers to an oligonucleotide, usually single-stranded, that provides a 3'-hydroxyl end for the initiation of enzyme-mediated nucleic acid synthesis. The following are non-limiting embodiments of polynucleotides: a gene or gene fragment, exons, introns, mRNA, tRNA, rRNA, ribozymes, cDNA, recombinant polynucleotides, branched polynucleotides, plasmids, vectors, isolated DNA of any sequence, isolated RNA of any sequence, nucleic acid probes and primers. A nucleic acid molecule may also comprise modified nucleic acid molecules, such as methylated nucleic acid molecules and nucleic acid molecule analogs. Analogs of purines and pyrimidines are known in the art, and include, but are not limited to, aziridinycytosine, 4-acetylcytosine, 5-fluorouracil, 5-bromouracil, 5-carboxymethylamino-methyl-2-thiouracil, 5-carboxymethyl-aminomethyluracil, inosine, N6-isopentenyladenine, 1-methyladenine, 1-methylpseudouracil, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, pseudouracil, 5-pentyluracil and 2,6-diaminopurine. The use of uracil as a substitute for thymine in a deoxyribonucleic acid is also considered an analogous form of pyrimidine.

15 A "fragment" of a polynucleotide is a polynucleotide comprised of at least 9 contiguous nucleotides, preferably at least 15 contiguous nucleotides and more preferably at least 45 nucleotides, of coding or non-coding sequences.

The term "gene targeting" refers to a type of homologous recombination that occurs when a fragment of genomic DNA is introduced into a mammalian cell and that fragment locates and recombines with endogenous homologous sequences.

20 The term "homologous recombination" refers to the exchange of DNA fragments between two DNA molecules or chromatids at the site of homologous nucleotide sequences.

The term "homologous" as used herein denotes a characteristic of a DNA sequence having at least about 70 percent sequence identity as compared to a reference sequence, typically at least about 25 85 percent sequence identity, preferably at least about 95 percent sequence identity, and more preferably about 98 percent sequence identity, and most preferably about 100 percent sequence identity as compared to a reference sequence. Homology can be determined using a "BLASTN" algorithm. It is understood that homologous sequences can accommodate insertions, deletions and substitutions in the nucleotide sequence. Thus, linear sequences of nucleotides can be essentially identical even if some of the nucleotide residues do not precisely correspond or align. The reference sequence may be a subset of a larger sequence, such as a portion of a gene or flanking sequence, or a repetitive portion of a chromosome.

30 The term "target gene" (alternatively referred to as "target gene sequence" or "target DNA sequence" or "target sequence") refers to any nucleic acid molecule or polynucleotide of any gene to be modified by homologous recombination. The target sequence includes an intact gene, an exon or intron, a regulatory sequence or any region between genes. The target gene comprises a portion of a

particular gene or genetic locus in the individual's genomic DNA. As provided herein, a "target gene" refers to: a down syndrome cell adhesion molecule-like (CHD2-like) gene, a thymic dendritic cell-derived factor 1 (TDCDF1) gene, a toll-like receptor 2 (TR2) gene, a ubiquitin ligase E3-like (E3-like) gene, a contrapsin-like serine protease inhibitor (CSPI) gene, or an ABC transporter ATPase 2 (Abcd2) gene.

A "down syndrome cell adhesion molecule-like gene" or a "CHD2-like gene" refers to a sequence comprising SEQ ID NO:1 or comprising the sequence encoding the down syndrome cell adhesion molecule (identified in Genbank as Accession No.: AA500290; GI NO: 2235257). In one aspect, the coding sequence of the down syndrome cell adhesion molecule-like gene comprises SEQ ID NO:1 or comprises the down syndrome cell adhesion molecule-like gene identified in Genbank as Accession No.: AA500290; GI NO: 2235257.

A "thymic dendritic cell-derived factor 1 gene" or a "TDCDF1 gene" refers to a sequence comprising SEQ ID NO:4 or comprising the sequence encoding the thymic dendritic cell-derived factor 1 (identified in Genbank as Accession No.: AF116911; GI NO: 6642776). In one aspect, the coding sequence of the thymic dendritic cell-derived factor 1 gene comprises SEQ ID NO:4 or comprises the thymic dendritic cell-derived factor 1 gene identified in Genbank as Accession No.: AF116911; GI NO: 6642776.

A "toll-like receptor 2 gene" or a "TR2 gene" refers to a sequence comprising SEQ ID NO:7 or comprising the sequence encoding the toll-like receptor 2 (identified in Genbank as Accession No.: AF216289; GI NO: 6760422). In one aspect, the coding sequence of the toll-like receptor 2 gene comprises SEQ ID NO:7 or comprises the toll-like receptor 2 gene identified in Genbank as Accession No.: AF216289; GI NO: 6760422.

A "ubiquitin ligase E3-like gene" or an "E3-like gene" refers to a sequence comprising SEQ ID NO:10 or comprising the sequence encoding the ubiquitin ligase E3 (identified in Genbank as Accession No.: AA497512; GI NO: 2232535). In one aspect, the coding sequence of the ubiquitin ligase E3 comprises SEQ ID NO:10 or comprises a ubiquitin ligase E3-like gene identified in Genbank as Accession No.: AA497512; GI NO: 2232535.

A "contrapsin-like serine protease inhibitor gene" or a "CSPI gene" refers to a sequence comprising SEQ ID NO:13 or comprising the sequence encoding the contrapsin-like serine protease inhibitor (identified in Genbank as Accession No.: W40642; GI NO: 1324958). In one aspect, the coding sequence of the contrapsin-like serine protease inhibitor SEQ ID NO:13 or comprises a contrapsin-like serine protease inhibitor gene identified in Genbank as Accession No.: W40642; GI NO: 1324958.

An "ABC transporter ATPase 2 gene" or an "Abcd2 gene" refers to a sequence comprising SEQ ID NO:16 or comprising the sequence encoding the ABC transporter ATPase 2 gene (identified in Genbank as Accession No.: NM011994; GI NO: 6752941). In one aspect, the coding sequence of

the ABC transporter ATPase 2 comprises SEQ ID NO:16 or comprises a ABC transporter ATPase 2 gene identified in Genbank as Accession No.: NM011994; GI NO: 6752941.

“Disruption” of a target gene occurs when a fragment of genomic DNA locates and recombines with an endogenous homologous sequence. These sequence disruptions or modifications may include insertions, missense, frameshift, deletion, or substitutions, or replacements of DNA sequence, or any combination thereof. Insertions include the insertion of entire genes, which may be of animal, plant, fungal, insect, prokaryotic, or viral origin. Disruption, for example, can alter or replace a promoter, enhancer, or splice site of a target gene, and can alter the normal gene product by inhibiting its production partially or completely or by enhancing the normal gene product’s activity.

The term, “transgenic cell”, refers to a cell containing within its genome a target gene that has been disrupted, modified, altered, or replaced completely or partially by the method of gene targeting.

The term “transgenic animal” refers to an animal that contains within its genome a specific gene that has been disrupted by the method of gene targeting. The transgenic animal includes both the heterozygote animal (*i.e.*, one defective allele and one wild-type allele) and the homozygous animal (*i.e.*, two defective alleles).

As used herein, the terms “selectable marker” or “positive selection marker” refers to a gene encoding a product that enables only the cells that carry the gene to survive and/or grow under certain conditions. For example, plant and animal cells that express the introduced neomycin resistance (*Neo^r*) gene are resistant to the compound G418. Cells that do not carry the *Neo^r* gene marker are killed by G418. Other positive selection markers will be known to those of skill in the art.

A “host cell” includes an individual cell or cell culture that can be or has been a recipient for vector(s) or for incorporation of nucleic acid molecules and/or proteins. Host cells include progeny of a single host cell, and the progeny may not necessarily be completely identical (in morphology or in total DNA complement) to the original parent due to natural, accidental, or deliberate mutation. A host cell includes cells transfected with the constructs of the present invention.

The term “modulates” as used herein refers to the inhibition, reduction, increase or enhancement of a target gene function, expression, activity, or alternatively a phenotype associated with a disruption in a target gene.

The term “ameliorates” refers to a decreasing, reducing or eliminating of a condition, disease, disorder, or phenotype, including an abnormality or symptom associated with a disruption in a targeted gene.

The term “abnormality” refers to any disease, disorder, condition, or phenotype in which a disruption of a targeted gene is implicated, including pathological conditions.

Brief Description of the Drawings

Figure 1 shows the polynucleotide sequence for a Down Syndrome Cell Adhesion Molecule-Like (CHD2-like) gene (SEQ ID NO:1).

Figure 2 (Panels A and B) shows design of the targeting construct used to disrupt the CHD2-like genes. Figure 2 (Panel B) shows the sequences identified as SEQ ID NO:2 and SEQ ID NO:3, which were used as the targeting arms (homologous sequences) in CHD2-like gene targeting construct.

5 Figure 3 shows the polynucleotide sequence for a Thymic Dendritic Cell-Derived Factor 1 (TDCDF1) gene (SEQ ID NO:4).

Figure 4 (Panels A and B) shows design of the targeting construct used to disrupt the TDCDF1 genes. Figure 4 (Panel B) shows the sequences identified as SEQ ID NO:5 and SEQ ID NO:6, which were used as the targeting arms (homologous sequences) in TDCDF1 gene targeting
10 construct.

Figure 5 shows the polynucleotide sequence for a Toll-Like Receptor 2 (TR2) gene (SEQ ID NO:7).

Figure 6 (Panels A and B) shows design of the targeting construct used to disrupt the TR2 genes. Figure 6 (Panel B) shows the sequences identified as SEQ ID NO:8 and SEQ ID NO:9, which
15 were used as the targeting arms (homologous sequences) in TR2 gene targeting construct.

Figure 7 shows the polynucleotide sequence for a Ubiquitin Ligase E3-Like (E3-like) gene (SEQ ID NO:10).

Figure 8 (Panels A and B) shows design of the targeting construct used to disrupt the E3-like genes. Figure 8 (Panel B) shows the sequences identified as SEQ ID NO:11 and SEQ ID NO:12,
20 which were used as the targeting arms (homologous sequences) in E3-like gene targeting construct.

Figure 9 shows the polynucleotide sequence for a Contrapsin-Like Serine Protease Inhibitor (CSPI) gene (SEQ ID NO:13).

Figure 10 (Panels A and B) shows design of the targeting construct used to disrupt the CSPI genes. Figure 10 (Panel B) shows the sequences identified as SEQ ID NO:14 and SEQ ID NO:15,
25 which were used as the targeting arms (homologous sequences) in CSPI gene targeting construct.

Figure 11 shows the polynucleotide sequence for a ABC Transporter ATPase 2 (Abcd2) gene (SEQ ID NO:16).

Figure 12 (Panels A and B) shows design of the targeting construct used to disrupt the Abcd2 genes. Figure 12 (Panel B) shows the sequences identified as SEQ ID NO:17 and SEQ ID NO:18,
30 which were used as the targeting arms (homologous sequences) in Abcd2 gene targeting construct.

Detailed Description of the Invention

The invention is based, in part, on the evaluation of the expression and role of genes and gene expression products, primarily those associated with a targeted gene. Among others, the invention permits the definition of disease pathways and the identification of diagnostically and therapeutically
35 useful targets. For example, genes that are mutated or down-regulated under disease conditions may be involved in causing or exacerbating the disease condition. Treatments directed at up-regulating the

activity of such genes or treatments that involve alternate pathways, may ameliorate the disease condition.

Generation of Targeting Construct

The targeting construct of the present invention may be produced using standard methods known in the art. (see, e.g., Sambrook, *et al.*, 1989, *Molecular Cloning: A Laboratory Manual*, Second Edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York; E.N. Glover (eds.), 1985, *DNA Cloning: A Practical Approach*, Volumes I and II; M.J. Gait (ed.), 1984, *Oligonucleotide Synthesis*; B.D. Hames & S.J. Higgins (eds.), 1985, *Nucleic Acid Hybridization*; B.D. Hames & S.J. Higgins (eds.), 1984, *Transcription and Translation*; R.I. Freshney (ed.), 1986, *Animal Cell Culture; Immobilized Cells and Enzymes*, IRL Press, 1986; B. Perbal, 1984, *A Practical Guide To Molecular Cloning*; F.M. Ausubel *et al.*, 1994, *Current Protocols in Molecular Biology*, John Wiley & Sons, Inc.). For example, the targeting construct may be prepared in accordance with conventional ways, where sequences may be synthesized, isolated from natural sources, manipulated, cloned, ligated, subjected to *in vitro* mutagenesis, primer repair, or the like. At various stages, the joined sequences may be cloned, and analyzed by restriction analysis, sequencing, or the like.

The targeting DNA can be constructed using techniques well known in the art. For example, the targeting DNA may be produced by chemical synthesis of oligonucleotides, nick-translation of a double-stranded DNA template, polymerase chain-reaction amplification of a sequence (or ligase chain reaction amplification), purification of prokaryotic or target cloning vectors harboring a sequence of interest (e.g., a cloned cDNA or genomic DNA, synthetic DNA or from any of the aforementioned combination) such as plasmids, phagemids, YACs, cosmids, bacteriophage DNA, other viral DNA or replication intermediates, or purified restriction fragments thereof, as well as other sources of single and double-stranded polynucleotides having a desired nucleotide sequence. Moreover, the length of homology may be selected using known methods in the art. For example, selection may be based on the sequence composition and complexity of the predetermined endogenous target DNA sequence(s).

The targeting construct of the present invention typically comprises a first sequence homologous to a portion or region of the targeted gene and a second sequence homologous to a second portion or region of the targeted gene. The targeting construct further comprises a positive selection marker, which is preferably positioned in between the first and the second DNA sequence that are homologous to a portion or region of the target DNA sequence. The positive selection marker may be operatively linked to a promoter and a polyadenylation signal.

Other regulatory sequences known in the art may be incorporated into the targeting construct to disrupt or control expression of a particular gene in a specific cell type. In addition, the targeting construct may also include a sequence coding for a screening marker, for example, green fluorescent protein (GFP), or another modified fluorescent protein.

Although the size of the homologous sequence is not critical and can range from as few as 50 base pairs to as many as 100 kb, preferably each fragment is greater than about 1 kb in length, more preferably between about 1 and about 10 kb, and even more preferably between about 1 and about 5 kb. One of skill in the art will recognize that although larger fragments may increase the number of homologous recombination events in ES cells, larger fragments will also be more difficult to clone.

In a preferred embodiment of the present invention, the targeting construct is prepared directly from a plasmid genomic library using the methods described in pending U.S. Patent Application Ser. No.: 08/971,310, filed November 17, 1997, the disclosure of which is incorporated herein in its entirety. Generally, a sequence of interest is identified and isolated from a plasmid library in a single step using, for example, long-range PCR. Following isolation of this sequence, a second polynucleotide that will disrupt the target sequence can be readily inserted between two regions encoding the sequence of interest. In accordance with this aspect, the construct is generated in two steps by (1) amplifying (for example, using long-range PCR) sequences homologous to the target sequence, and (2) inserting another polynucleotide (for example a selectable marker) into the PCR product so that it is flanked by the homologous sequences. Typically, the vector is a plasmid from a plasmid genomic library. The completed construct is also typically a circular plasmid.

In another embodiment, the targeting construct is designed in accordance with the regulated positive selection method described in U.S. Patent Application Ser. No. 60/232,957, filed September 15, 2000, the disclosure of which is incorporated herein in its entirety. The targeting construct is designed to include a PGK-*neo* fusion gene having two *lacO* sites, positioned in the PGK promoter and an NLS-*lacI* gene comprising a lac repressor fused to sequences encoding the NLS from the SV40 T antigen.

In another embodiment, the targeting construct may contain more than one selectable marker gene, including a negative selectable marker, such as the herpes simplex virus tk (HSV-tk) gene. The negative selectable marker may be operatively linked to a promoter and a polyadenylation signal. (*see, e.g.*, U.S. Patent No. 5,464,764; U.S. Patent No. 5,487,992; U.S. Patent No. 5,627,059; and U.S. Patent No. 5,631,153).

Generation of Cells and Confirmation of Homologous Recombination Events

Once an appropriate targeting construct has been prepared, the targeting construct may be introduced into an appropriate host cell using any method known in the art. Various techniques may be employed in the present invention, including, for example, pronuclear microinjection; retrovirus mediated gene transfer into germ lines; gene targeting in embryonic stem cells; electroporation of embryos; sperm-mediated gene transfer; and calcium phosphate/DNA co-precipitates, microinjection of DNA into the nucleus, bacterial protoplast fusion with intact cells, transfection, polycations, *e.g.*, polybrene, polyornithine, *etc.*, or the like (*see, e.g.*, U.S. Patent No. 4,873,191; Van der Putten, *et al.*, 1985, *Proc. Natl. Acad. Sci., USA* 82:6148-6152; Thompson, *et al.*, 1989, *Cell* 56:313-321; Lo, 1983, *Mol Cell. Biol.* 3:1803-1814; Lavitrano, *et al.*, 1989, *Cell*, 57:717-723). Various techniques for

transforming mammalian cells are known in the art (*see, e.g.,* Gordon, 1989, *Intl. Rev. Cytol.*, 115:171-229; Keown *et al.*, 1989, *Methods in Enzymology*; Keown *et al.*, 1990, *Methods and Enzymology*, Vol. 185, pp. 527-537; Mansour *et al.*, 1988, *Nature*, 336:348-352).

5 In a preferred aspect of the present invention, the targeting construct is introduced into host cells by electroporation. In this process, electrical impulses of high field strength reversibly permeabilize biomembranes allowing the introduction of the construct. The pores created during electroporation permit the uptake of macromolecules such as DNA (*see, e.g.,* Potter, H., *et al.*, 1984, *Proc. Nat'l. Acad. Sci. U.S.A.* 81:7161-7165).

10 Any cell type capable of homologous recombination may be used in the practice of the present invention. Examples of such target cells include cells derived from vertebrates including mammals such as humans, bovine species, ovine species, murine species, simian species, and other eucaryotic organisms such as filamentous fungi, and higher multicellular organisms such as plants.

Preferred cell types include embryonic stem (ES) cells, which are typically obtained from pre-implantation embryos cultured *in vitro* (*see, e.g.,* Evans, M. J., *et al.*, 1981, *Nature* 292:154-156; 15 Bradley, M. O., *et al.*, 1984, *Nature* 309:255-258; Gossler *et al.*, 1986, *Proc. Natl. Acad. Sci. USA* 83:9065-9069; and Robertson, *et al.*, 1986, *Nature* 322:445-448). The ES cells are cultured and prepared for introduction of the targeting construct using methods well known to the skilled artisan (*see, e.g.,* Robertson, E. J. ed. "Teratocarcinomas and Embryonic Stem Cells, a Practical Approach", IRL Press, Washington D.C., 1987; Bradley *et al.*, 1986, *Current Topics in Devel. Biol.* 20:357-371; 20 by Hogan *et al.*, in "Manipulating the Mouse Embryo": A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor N.Y., 1986; Thomas *et al.*, 1987, *Cell* 51:503; Koller *et al.*, 1991, *Proc. Natl. Acad. Sci. USA*, 88:10730; Dorin *et al.*, 1992, *Transgenic Res.* 1:101; and Veis *et al.*, 1993, *Cell* 75:229). The ES cells that will be inserted with the targeting construct are derived from an embryo or blastocyst of the same species as the developing embryo into which they are to be 25 introduced. ES cells are typically selected for their ability to integrate into the inner cell mass and contribute to the germ line of an individual when introduced into the mammal in an embryo at the blastocyst stage of development. Thus, any ES cell line having this capability is suitable for use in the practice of the present invention.

The present invention may also be used to knockout genes in other cell types, such as stem 30 cells. By way of example, stem cells may be myeloid, lymphoid, or neural progenitor and precursor cells. These cells comprising a disruption or knockout of a gene may be particularly useful in the study of targeted gene function in individual developmental pathways. Stem cells may be derived from any vertebrate species, such as mouse, rat, dog, cat, pig, rabbit, human, non-human primates and the like.

35 After the targeting construct has been introduced into cells, the cells where successful gene targeting has occurred are identified. Insertion of the targeting construct into the targeted gene is

typically detected by identifying cells for expression of the marker gene. In a preferred embodiment, the cells transformed with the targeting construct of the present invention are subjected to treatment with an appropriate agent that selects against cells not expressing the selectable marker. Only those cells expressing the selectable marker gene survive and/or grow under certain conditions. For
5 example, cells that express the introduced neomycin resistance gene are resistant to the compound G418, while cells that do not express the neo gene marker are killed by G418. If the targeting construct also comprises a screening marker such as GFP, homologous recombination can be identified through screening cell colonies under a fluorescent light. Cells that have undergone homologous recombination will have deleted the GFP gene and will not fluoresce.

10 If a regulated positive selection method is used in identifying homologous recombination events, the targeting construct is designed so that the expression of the selectable marker gene is regulated in a manner such that expression is inhibited following random integration but is permitted (derepressed) following homologous recombination. More particularly, the transfected cells are screened for expression of the *neo* gene, which requires that (1) the cell was successfully
15 electroporated, and (2) *lac* repressor inhibition of *neo* transcription was relieved by homologous recombination. This method allows for the identification of transfected cells and homologous recombinants to occur in one step with the addition of a single drug.

Alternatively, a positive-negative selection technique may be used to select homologous recombinants. This technique involves a process in which a first drug is added to the cell population,
20 for example, a neomycin-like drug to select for growth of transfected cells, *i.e.* positive selection. A second drug, such as FIAU is subsequently added to kill cells that express the negative selection marker, *i.e.* negative selection. Cells that contain and express the negative selection marker are killed by a selecting agent, whereas cells that do not contain and express the negative selection marker survive. For example, cells with non-homologous insertion of the construct express HSV thymidine
25 kinase and therefore are sensitive to the herpes drugs such as gancyclovir (GANC) or FIAU (1-(2-deoxy 2-fluoro-B-D-arabinofluranosyl)-5-iodouracil) (*see, e.g., Mansour et al., Nature* 336:348-352: (1988); Capecchi, *Science* 244:1288-1292, (1989); Capecchi, *Trends in Genet.* 5:70-76 (1989)).

Successful recombination may be identified by analyzing the DNA of the selected cells to confirm homologous recombination. Various techniques known in the art, such as PCR and/or
30 Southern analysis may be used to confirm homologous recombination events.

Homologous recombination may also be used to disrupt genes in stem cells, and other cell types, which are not totipotent embryonic stem cells. By way of example, stem cells may be myeloid, lymphoid, or neural progenitor and precursor cells. Such transgenic cells may be particularly useful in the study of targeted gene function in individual developmental pathways. Stem cells may be
35 derived from any vertebrate species, such as mouse, rat, dog, cat, pig, rabbit, human, non-human primates and the like.

In cells that are not totipotent it may be desirable to knock out both copies of the target using methods that are known in the art. For example, cells comprising homologous recombination at a target locus that have been selected for expression of a positive selection marker (*e.g.*, *Neo*^r) and screened for non-random integration, can be further selected for multiple copies of the selectable marker gene by exposure to elevated levels of the selective agent (*e.g.*, G418). The cells are then analyzed for homozygosity at the target locus. Alternatively, a second construct can be generated with a different positive selection marker inserted between the two homologous sequences. The two constructs can be introduced into the cell either sequentially or simultaneously, followed by appropriate selection for each of the positive marker genes. The final cell is screened for homologous recombination of both alleles of the target.

Production of Transgenic Animals

Selected cells are then injected into a blastocyst (or other stage of development suitable for the purposes of creating a viable animal, such as, for example, a morula) of an animal (*e.g.*, a mouse) to form chimeras (*see e.g.*, Bradley, A. in *Teratocarcinomas and Embryonic Stem Cells: A Practical Approach*, E. J. Robertson, ed., IRL, Oxford, pp. 113-152 (1987)). Alternatively, selected ES cells can be allowed to aggregate with dissociated mouse embryo cells to form the aggregation chimera. A chimeric embryo can then be implanted into a suitable pseudopregnant female foster animal and the embryo brought to term. Chimeric progeny harboring the homologously recombined DNA in their germ cells can be used to breed animals in which all cells of the animal contain the homologously recombined DNA. In one embodiment, chimeric progeny mice are used to generate a mouse with a heterozygous disruption in the targeted gene. Heterozygous transgenic mice can then be mated. It is well known in the art that typically 1/4 of the offspring of such matings will have a homozygous disruption in the targeted gene.

The heterozygous and homozygous transgenic mice can then be compared to normal, wild type mice to determine whether disruption of the targeted gene causes phenotypic changes, especially pathological changes. For example, heterozygous and homozygous mice may be evaluated for phenotypic changes by physical examination, necropsy, histology, clinical chemistry, complete blood count, body weight, organ weights, and cytological evaluation of bone marrow.

In one embodiment, the phenotype (or phenotypic change) associated with a disruption in the targeted gene is placed into or stored in a database. Preferably, the database includes: (i) genotypic data (*e.g.*, identification of the disrupted gene) and (ii) phenotypic data (*e.g.*, phenotype(s) resulting from the gene disruption) associated with the genotypic data. The database is preferably electronic. In addition, the database is preferably combined with a search tool so that the database is searchable.

Conditional Transgenic Animals

The present invention further contemplates conditional transgenic or knockout animals, such as those produced using recombination methods. Bacteriophage P1 Cre recombinase and flp

recombinase from yeast plasmids are two non-limiting examples of site-specific DNA recombinase enzymes that cleave DNA at specific target sites (lox P sites for cre recombinase and frt sites for flp recombinase) and catalyze a ligation of this DNA to a second cleaved site. A large number of suitable alternative site-specific recombinases have been described, and their genes can be used in accordance with the method of the present invention. Such recombinases include the Int recombinase of bacteriophage λ (with or without Xis) (Weisberg, R. *et al.*, in *Lambda II*, (Hendrix, R., *et al.*, Eds.), Cold Spring Harbor Press, Cold Spring Harbor, NY, pp. 211-50 (1983), herein incorporated by reference); TpnI and the β -lactamase transposons (Mercier, *et al.*, *J. Bacteriol.*, 172:3745-57 (1990)); the Tn3 resolvase (Flanagan & Fennewald *J. Molec. Biol.*, 206:295-304 (1989); Stark, *et al.*, *Cell*, 58:779-90 (1989)); the yeast recombinases (Matsuzaki, *et al.*, *J. Bacteriol.*, 172:610-18 (1990)); the *B. subtilis* SpoIVC recombinase (Sato, *et al.*, *J. Bacteriol.* 172:1092-98 (1990)); the Flp recombinase (Schwartz & Sadowski, *J. Molec. Biol.*, 205:647-658 (1989); Parsons, *et al.*, *J. Biol. Chem.*, 265:4527-33 (1990); Golic & Lindquist, *Cell*, 59:499-509 (1989); Amin, *et al.*, *J. Molec. Biol.*, 214:55-72 (1990)); the Hin recombinase (Glasgow, *et al.*, *J. Biol. Chem.*, 264:10072-82 (1989)); immunoglobulin recombinases (Malynn, *et al.*, *Cell*, 54:453-460 (1988)); and the Cin recombinase (Haffter & Bickle, *EMBO J.*, 7:3991-3996 (1988); Hubner, *et al.*, *J. Molec. Biol.*, 205:493-500 (1989)), all herein incorporated by reference. Such systems are discussed by Echols (*J. Biol. Chem.* 265:14697-14700 (1990)); de Villartay (*Nature*, 335:170-74 (1988)); Craig, (*Ann. Rev. Genet.*, 22:77-105 (1988)); Poyart-Salmeron, *et al.*, (*EMBO J.* 8:2425-33 (1989)); Hunger-Bertling, *et al.*, (*Mol. Cell. Biochem.*, 92:107-16 (1990)); and Cregg & Madden (*Mol. Gen. Genet.*, 219:320-23 (1989)), all herein incorporated by reference.

Cre has been purified to homogeneity, and its reaction with the loxP site has been extensively characterized (Abremski & Hess *J. Mol. Biol.* 259:1509-14 (1984), herein incorporated by reference). Cre protein has a molecular weight of 35,000 and can be obtained commercially from New England Nuclear/DuPont. The cre gene (which encodes the Cre protein) has been cloned and expressed (Abremski, *et al.*, *Cell* 32:1301-11 (1983), herein incorporated by reference). The Cre protein mediates recombination between two loxP sequences (Sternberg, *et al.*, *Cold Spring Harbor Symp. Quant. Biol.* 45:297-309 (1981)), which may be present on the same or different DNA molecule. Because the internal spacer sequence of the loxP site is asymmetrical, two loxP sites can exhibit directionality relative to one another (Hoess & Abremski *Proc. Natl. Acad. Sci. U.S.A.* 81:1026-29 (1984)). Thus, when two sites on the same DNA molecule are in a directly repeated orientation, Cre will excise the DNA between the sites (Abremski, *et al.*, *Cell* 32:1301-11 (1983)). However, if the sites are inverted with respect to each other, the DNA between them is not excised after recombination but is simply inverted. Thus, a circular DNA molecule having two loxP sites in direct orientation will recombine to produce two smaller circles, whereas circular molecules having two loxP sites in an inverted orientation simply invert the DNA sequences flanked by the loxP sites. In

addition, recombinase action can result in reciprocal exchange of regions distal to the target site when targets are present on separate DNA molecules.

Recombinases have important application for characterizing gene function in knockout models. When the constructs described herein are used to disrupt targeted genes, a fusion transcript
5 can be produced when insertion of the positive selection marker occurs downstream (3') of the translation initiation site of the targeted gene. The fusion transcript could result in some level of protein expression with unknown consequence. It has been suggested that insertion of a positive selection marker gene can affect the expression of nearby genes. These effects may make it difficult to determine gene function after a knockout event since one could not discern whether a given
10 phenotype is associated with the inactivation of a gene, or the transcription of nearby genes. Both potential problems are solved by exploiting recombinase activity. When the positive selection marker is flanked by recombinase sites in the same orientation, the addition of the corresponding recombinase will result in the removal of the positive selection marker. In this way, effects caused by the positive selection marker or expression of fusion transcripts are avoided.

15 In one embodiment, purified recombinase enzyme is provided to the cell by direct microinjection. In another embodiment, recombinase is expressed from a co-transfected construct or vector in which the recombinase gene is operably linked to a functional promoter. An additional aspect of this embodiment is the use of tissue-specific or inducible recombinase constructs that allow the choice of when and where recombination occurs. One method for practicing the inducible forms
20 of recombinase-mediated recombination involves the use of vectors that use inducible or tissue-specific promoters or other gene regulatory elements to express the desired recombinase activity. The inducible expression elements are preferably operatively positioned to allow the inducible control or activation of expression of the desired recombinase activity. Examples of such inducible promoters or other gene regulatory elements include, but are not limited to, tetracycline, metallothionine, ecdysone,
25 and other steroid-responsive promoters, rapamycin responsive promoters, and the like (No, *et al.*, *Proc. Natl. Acad. Sci. USA*, 93:3346-51 (1996); Furth, *et al.*, *Proc. Natl. Acad. Sci. USA*, 91:9302-6 (1994)). Additional control elements that can be used include promoters requiring specific transcription factors such as viral, promoters. Vectors incorporating such promoters would only express recombinase activity in cells that express the necessary transcription factors.

30 Models for Disease

The cell- and animal-based systems described herein can be utilized as models for diseases. Animals of any species, including, but not limited to, mice, rats, rabbits, guinea pigs, pigs, micro-pigs, goats, and non-human primates, *e.g.*, baboons, monkeys, and chimpanzees may be used to generate disease animal models. In addition, cells from humans may be used. These systems may be used in a
35 variety of applications. Such assays may be utilized as part of screening strategies designed to identify agents, such as compounds that are capable of ameliorating disease symptoms. Thus, the

animal- and cell-based models may be used to identify drugs, pharmaceuticals, therapies and interventions that may be effective in treating disease.

Cell-based systems may be used to identify compounds that may act to ameliorate disease symptoms. For example, such cell systems may be exposed to a compound suspected of exhibiting an ability to ameliorate disease symptoms, at a sufficient concentration and for a time sufficient to elicit such an amelioration of disease symptoms in the exposed cells. After exposure, the cells are examined to determine whether one or more of the disease cellular phenotypes has been altered to resemble a more normal or more wild type, non-disease phenotype.

In addition, animal-based disease systems, such as those described herein, may be used to identify compounds capable of ameliorating disease symptoms. Such animal models may be used as test substrates for the identification of drugs, pharmaceuticals, therapies, and interventions that may be effective in treating a disease or other phenotypic characteristic of the animal. For example, animal models may be exposed to a compound or agent suspected of exhibiting an ability to ameliorate disease symptoms, at a sufficient concentration and for a time sufficient to elicit such an amelioration of disease symptoms in the exposed animals. The response of the animals to the exposure may be monitored by assessing the reversal of disorders associated with the disease. Exposure may involve treating mother animals during gestation of the model animals described herein, thereby exposing embryos or fetuses to the compound or agent that may prevent or ameliorate the disease or phenotype. Neonatal, juvenile, and adult animals can also be exposed.

More particularly, using the animal models of the invention, specifically, transgenic mice, methods of identifying agents, including compounds are provided, preferably, on the basis of the ability to affect at least one phenotype associated with a disruption in a targeted gene. In one embodiment, the present invention provides a method of identifying agents having an effect on target gene expression or function. The method includes measuring a physiological response of the animal, for example, to the agent, and comparing the physiological response of such animal to a control animal, wherein the physiological response of the animal comprising a disruption in a target gene as compared to the control animal indicates the specificity of the agent. A "physiological response" is any biological or physical parameter of an animal that can be measured. Molecular assays (*e.g.*, gene transcription, protein production and degradation rates), physical parameters (*e.g.*, exercise physiology tests, measurement of various parameters of respiration, measurement of heart rate or blood pressure, measurement of bleeding time, aPTT, or TT), and cellular assays (*e.g.*, immunohistochemical assays of cell surface markers, or the ability of cells to aggregate or proliferate) can be used to assess a physiological response.

The transgenic animals and cells of the present invention may be utilized as models for diseases, disorders, or conditions associated with phenotypes relating to a disruption in a target gene. The present invention also provides a unique animal model for testing and developing new treatments

relating to the behavioral phenotypes. Analysis of the behavioral phenotype allows for the development of an animal model useful for testing, for instance, the efficacy of proposed genetic and pharmacological therapies for human genetic diseases, such as neurological, neuropsychological, or psychotic illnesses.

5 A statistical analysis of the various behaviors measured can be carried out using any conventional statistical program routinely used by those skilled in the art (such as, for example, "Analysis of Variance" or ANOVA). A "p" value of about 0.05 or less is generally considered to be statistically significant, although slightly higher p values may still be indicative of statistically significant differences. To statistically analyze abnormal behavior, a comparison is made between the
10 behavior of a transgenic animal (or a group thereof) to the behavior of a wild-type mouse (or a group thereof), typically under certain prescribed conditions. "Abnormal behavior" as used herein refers to behavior exhibited by an animal having a disruption in the targeted gene, *e.g.* transgenic animal, which differs from an animal without a disruption in the targeted gene, *e.g.* wild-type mouse. Abnormal behavior consists of any number of standard behaviors that can be objectively measured (or
15 observed) and compared. In the case of comparison, it is preferred that the change be statistically significant to confirm that there is indeed a meaningful behavioral difference between the knockout animal and the wild-type control animal. Examples of behaviors that may be measured or observed include, but are not limited to, ataxia, rapid limb movement, eye movement, breathing, motor activity, cognition, emotional behaviors, social behaviors, hyperactivity, hypersensitivity, anxiety, impaired
20 learning, abnormal reward behavior, and abnormal social interaction, such as aggression.

A series of tests may be used to measure the behavioral phenotype of the animal models of the present invention, including neurological and neuropsychological tests to identify abnormal behavior. These tests may be used to measure abnormal behavior relating to, for example, learning and memory, eating, pain, aggression, sexual reproduction, anxiety, depression, schizophrenia, and drug abuse. (*see, e.g.,* Crawley & Paylor, *Hormones and Behavior* 31:197-211 (1997)).
25

The social interaction test involves exposing a mouse to other animals in a variety of settings. The social behaviors of the animals (*e.g.,* touching, climbing, sniffing, and mating) are subsequently evaluated. Differences in behaviors can then be statistically analyzed and compared (*see, e.g.,* S. E. File, *et al., Pharmacol. Bioch. Behav.* 22:941-944 (1985); R. R. Holson, *Phys. Behav.* 37:239-247
30 (1986)). Exemplary behavioral tests include the following.

The mouse startle response test typically involves exposing the animal to a sensory (typically auditory) stimulus and measuring the startle response of the animal (*see, e.g.,* M. A. Geyer, *et al., Brain Res. Bull.* 25:485-498 (1990); Paylor and Crawley, *Psychopharmacology* 132:169-180 (1997)). A pre-pulse inhibition test can also be used, in which the percent inhibition (from a normal startle
35 response) is measured by "cueing" the animal first with a brief low-intensity pre-pulse prior to the startle pulse.

The electric shock test generally involves exposure to an electrified surface and measurement of subsequent behaviors such as, for example, motor activity, learning, social behaviors. The behaviors are measured and statistically analyzed using standard statistical tests. (*see, e.g., G. J. Kant, et al., Pharm. Bioch. Behav.* 20:793-797 (1984); N. J. Leidenheimer, *et al., Pharmacol. Bioch. Behav.* 30:351-355 (1988)).

The tail-pinch or immobilization test involves applying pressure to the tail of the animal and/or restraining the animal's movements. Motor activity, social behavior, and cognitive behavior are examples of the areas that are measured. (*see, e.g., M. Bertolucci D'Angic, et al., Neurochem.* 55:1208-1214 (1990)).

The novelty test generally comprises exposure to a novel environment and/or novel objects. The animal's motor behavior in the novel environment and/or around the novel object are measured and statistically analyzed. (*see, e.g., D. K. Reinstein, et al., Pharm. Bioch. Behav.* 17:193-202 (1982); B. Poucet, *Behav. Neurosci.* 103:1009-10016 (1989); R. R. Holson, *et al., Phys. Behav.* 37:231-238 (1986)). This test may be used to detect visual processing deficiencies or defects.

The learned helplessness test involves exposure to stresses, for example, noxious stimuli, which cannot be affected by the animal's behavior. The animal's behavior can be statistically analyzed using various standard statistical tests. (*see, e.g., A. Leshner, et al., Behav. Neural Biol.* 26:497-501 (1979)).

Alternatively, a tail suspension test may be used, in which the "immobile" time of the mouse is measured when suspended "upside-down" by its tail. This is a measure of whether the animal struggles, an indicator of depression. In humans, depression is believed to result from feelings of a lack of control over one's life or situation. It is believed that a depressive state can be elicited in animals by repeatedly subjecting them to aversive situations over which they have no control. A condition of "learned helplessness" is eventually reached, in which the animal will stop trying to change its circumstances and simply accept its fate. Animals that stop struggling sooner are believed to be more prone to depression. Studies have shown that the administration of certain antidepressant drugs prior to testing increases the amount of time that animals struggle before giving up.

The Morris water-maze test comprises learning spatial orientations in water and subsequently measuring the animal's behaviors, such as, for example, by counting the number of incorrect choices. The behaviors measured are statistically analyzed using standard statistical tests. (*see, e.g., E. M. Spruijt, et al., Brain Res.* 527:192-197 (1990)).

Alternatively, a Y-shaped maze may be used (*see, e.g., McFarland, D.J., Pharmacology, Biochemistry and Behavior* 32:723-726 (1989); Dellu, F., *et al., Neurobiology of Learning and Memory* 73:31-48 (2000)). The Y-maze is generally believed to be a test of cognitive ability. The dimensions of each arm of the Y-maze can be, for example, approximately 40 cm x 8 cm x 20 cm, although other dimensions may be used. Each arm can also have, for example, sixteen equally spaced

photobeams to automatically detect movement within the arms. At least two different tests can be performed using such a Y-maze. In a continuous Y-maze paradigm, mice are allowed to explore all three arms of a Y-maze for, *e.g.*, approximately 10 minutes. The animals are continuously tracked using photobeam detection grids, and the data can be used to measure spontaneous alteration and positive bias behavior. Spontaneous alteration refers to the natural tendency of a "normal" animal to visit the least familiar arm of a maze. An alternation is scored when the animal makes two consecutive turns in the same direction, thus representing a sequence of visits to the least recently entered arm of the maze. Position bias determines egocentrically defined responses by measuring the animal's tendency to favor turning in one direction over another. Therefore, the test can detect differences in an animal's ability to navigate on the basis of allocentric or egocentric mechanisms. The two-trial Y-maze memory test measures response to novelty and spatial memory based on a free-choice exploration paradigm. During the first trial (acquisition), the animals are allowed to freely visit two arms of the Y-maze for, *e.g.*, approximately 15 minutes. The third arm is blocked off during this trial. The second trial (retrieval) is performed after an intertrial interval of, *e.g.*, approximately 2 hours. During the retrieval trial, the blocked arm is opened and the animal is allowed access to all three arms for, *e.g.*, approximately 5 minutes. Data are collected during the retrieval trial and analyzed for the number and duration of visits to each arm. Because the three arms of the maze are virtually identical, discrimination between novelty and familiarity is dependent on "environmental" spatial cues around the room relative to the position of each arm. Changes in arm entry and duration of time spent in the novel arm in a transgenic animal model may be indicative of a role of that gene in mediating novelty and recognition processes.

The passive avoidance or shuttle box test generally involves exposure to two or more environments, one of which is noxious, providing a choice to be learned by the animal. Behavioral measures include, for example, response latency, number of correct responses, and consistency of response. (*see, e.g.*, R. Ader, *et al.*, *Psychon. Sci.* 26:125-128 (1972); R. R. Holson, *Phys. Behav.* 37:221-230 (1986)). Alternatively, a zero-maze can be used. In a zero-maze, the animals can, for example, be placed in a closed quadrant of an elevated annular platform having, *e.g.*, 2 open and 2 closed quadrants, and are allowed to explore for approximately 5 minutes. This paradigm exploits an approach-avoidance conflict between normal exploratory activity and an aversion to open spaces in rodents. This test measures anxiety levels and can be used to evaluate the effectiveness of anti-anxiolytic drugs. The time spent in open quadrants versus closed quadrants may be recorded automatically, with, for example, the placement of photobeams at each transition site.

The food avoidance test involves exposure to novel food and objectively measuring, for example, food intake and intake latency. The behaviors measured are statistically analyzed using standard statistical tests. (*see, e.g.*, B. A. Campbell, *et al.*, *J. Comp. Physiol. Psychol.* 67:15-22 (1969)).

The elevated plus-maze test comprises exposure to a maze, without sides, on a platform, the animal's behavior is objectively measured by counting the number of maze entries and maze learning. The behavior is statistically analyzed using standard statistical tests. (*see, e.g., H. A. Baldwin, et al., Brain Res. Bull, 20:603-606 (1988)*).

5 The stimulant-induced hyperactivity test involves injection of stimulant drugs (*e.g., amphetamines, cocaine, PCP, and the like*), and objectively measuring, for example, motor activity, social interactions, cognitive behavior. The animal's behaviors are statistically analyzed using standard statistical tests. (*see, e.g., P. B. S. Clarke, et al., Psychopharmacology 96:511-520 (1988); P. Kuczenski, et al., J. Neuroscience 11:2703-2712 (1991)*).

10 The self-stimulation test generally comprises providing the mouse with the opportunity to regulate electrical and/or chemical stimuli to its own brain. Behavior is measured by frequency and pattern of self-stimulation. Such behaviors are statistically analyzed using standard statistical tests. (*see, e.g., S. Nassif, et al., Brain Res., 332:247-257 (1985); W. L. Isaac, et al., Behav. Neurosci. 103:345-355 (1989)*).

15 The reward test involves shaping a variety of behaviors, *e.g., motor, cognitive, and social*, measuring, for example, rapidity and reliability of behavioral change, and statistically analyzing the behaviors measured. (*see, e.g., L. E. Jarrard, et al., Exp. Brain Res. 61:519-530 (1986)*).

 The DRL (differential reinforcement to low rates of responding) performance test involves exposure to intermittent reward paradigms and measuring the number of proper responses, *e.g., lever pressing*. Such behavior is statistically analyzed using standard statistical tests. (*see, e.g., J. D. Sinden, et al., Behav. Neurosci. 100:320-329 (1986); V. Nalwa, et al., Behav Brain Res. 17:73-76 (1985); and A. J. Nonneman, et al., J. Comp. Physiol. Psych. 95:588-602 (1981)*).

 The spatial learning test involves exposure to a complex novel environment, measuring the rapidity and extent of spatial learning, and statistically analyzing the behaviors measured. (*see, e.g.,*
25 N. Pitsikas, *et al., Pharm. Bioch. Behav. 38:931-934 (1991); B. Poucet, et al., Brain Res. 37:269-280 (1990); D. Christie, et al., Brain Res. 37:263-268 (1990); and F. Van Haaren, et al., Behav. Neurosci. 102:481-488 (1988)*). Alternatively, an open-field (of) test may be used, in which the greater distance traveled for a given amount of time is a measure of the activity level and anxiety of the animal. When the open field is a novel environment, it is believed that an approach-avoidance situation is created, in
30 which the animal is "torn" between the drive to explore and the drive to protect itself. Because the chamber is lighted and has no places to hide other than the corners, it is expected that a "normal" mouse will spend more time in the corners and around the periphery than it will in the center where there is no place to hide. "Normal" mice will, however, venture into the central regions as they explore more and more of the chamber. It can then be extrapolated that especially anxious mice will
35 spend most of their time in the corners, with relatively little or no exploration of the central region,

whereas bold (*i.e.*, less anxious) mice will travel a greater distance, showing little preference for the periphery versus the central region.

The visual, somatosensory and auditory neglect tests generally comprise exposure to a sensory stimulus, objectively measuring, for example, orientating responses, and statistically
5 analyzing the behaviors measured. (*see, e.g.*, J. M. Vargo, *et al.*, *Exp. Neurol.* 102:199-209 (1988)).

The consummatory behavior test generally comprises feeding and drinking, and objectively measuring quantity of consumption. The behavior measured is statistically analyzed using standard statistical tests. (*see, e.g.*, P. J. Fletcher, *et al.*, *Psychopharmacol.* 102:301-308 (1990); M. G. Corda, *et al.*, *Proc. Nat'l Acad. Sci. USA* 80:2072-2076 (1983)).

10 A visual discrimination test can also be used to evaluate the visual processing of an animal. One or two similar objects are placed in an open field and the animal is allowed to explore for about 5-10 minutes. The time spent exploring each object (proximity to, *i.e.*, movement within, *e.g.*, about 3-5 cm of the object is considered exploration of an object) is recorded. The animal is then removed from the open field, and the objects are replaced by a similar object and a novel object. The animal is
15 returned to the open field and the percent time spent exploring the novel object over the old object is measured (again, over about a 5-10 minute span). "Normal" animals will typically spend a higher percentage of time exploring the novel object rather than the old object. If a delay is imposed between sampling and testing, the memory task becomes more hippocampal-dependent. If no delay is imposed, the task is more based on simple visual discrimination. This test can also be used for
20 olfactory discrimination, in which the objects (preferably, simple blocks) can be sprayed or otherwise treated to hold an odor. This test can also be used to determine if the animal can make gustatory discriminations; animals that return to the previously eaten food instead of novel food exhibit gustatory neophobia.

A hot plate analgesia test can be used to evaluate an animal's sensitivity to heat or painful
25 stimuli. For example, a mouse can be placed on an approximately 55°C hot plate and the mouse's response latency (*e.g.*, time to pick up and lick a hind paw) can be recorded. These responses are not reflexes, but rather "higher" responses requiring cortical involvement. This test may be used to evaluate a nociceptive disorder.

An accelerating rotarod test may be used to measure coordination and balance in mice.
30 Animals can be, for example, placed on a rod that acts like a rotating treadmill (or rolling log). The rotarod can be made to rotate slowly at first and then progressively faster until it reaches a speed of, *e.g.*, approximately 60 rpm. The mice must continually reposition themselves in order to avoid falling off. The animals are preferably tested in at least three trials, a minimum of 20 minutes apart. Those mice that are able to stay on the rod the longest are believed to have better coordination and balance.

35 A metrazol administration test can be used to screen animals for varying susceptibilities to seizures or similar events. For example, a 5mg/ml solution of metrazol can be infused through the tail

vein of a mouse at a rate of, *e.g.*, approximately 0.375 ml/min. The infusion will cause all mice to experience seizures, followed by death. Those mice that enter the seizure stage the soonest are believed to be more prone to seizures. Four distinct physiological stages can be recorded: soon after the start of infusion, the mice will exhibit a noticeable "twitch", followed by a series of seizures,
5 ending in a final tensing of the body known as "tonic extension", which is followed by death.

Target Gene Products

The present invention further contemplates use of the target gene sequence to produce target gene products. Target gene products may include proteins that represent functionally equivalent gene products. Such an equivalent gene product may contain deletions, additions or substitutions of amino
10 acid residues within the amino acid sequence encoded by the gene sequences described herein, but which result in a silent change, thus producing a functionally equivalent target gene product. Amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues involved.

For example, nonpolar (hydrophobic) amino acids include alanine, leucine, isoleucine, valine,
15 proline, phenylalanine, tryptophan, and methionine; polar neutral amino acids include glycine, serine, threonine, cysteine, tyrosine, asparagine, and glutamine; positively charged (basic) amino acids include arginine, lysine, and histidine; and negatively charged (acidic) amino acids include aspartic acid and glutamic acid. "Functionally equivalent", as utilized herein, refers to a protein capable of exhibiting a substantially similar *in vivo* activity as the endogenous gene products encoded by the
20 targeted gene sequences. Alternatively, when utilized as part of an assay, "functionally equivalent" may refer to peptides capable of interacting with other cellular or extracellular molecules in a manner substantially similar to the way in which the corresponding portion of the endogenous gene product would.

Other protein products useful according to the methods of the invention are peptides derived
25 from or based on the target gene produced by recombinant or synthetic means (derived peptides).

Target gene products may be produced by recombinant DNA technology using techniques well known in the art. Thus, methods for preparing the gene polypeptides and peptides of the invention by expressing nucleic acid encoding gene sequences are described herein. Methods that are well known to those skilled in the art can be used to construct expression vectors containing gene
30 protein coding sequences and appropriate transcriptional/translational control signals. These methods include, for example, *in vitro* recombinant DNA techniques, synthetic techniques and *in vivo* recombination/genetic recombination (*see, e.g.*, Sambrook, *et al.*, 1989, *supra*, and Ausubel, *et al.*, 1989, *supra*). Alternatively, RNA capable of encoding gene protein sequences may be chemically synthesized using, for example, automated synthesizers (*see, e.g.* Oligonucleotide Synthesis: A
35 Practical Approach, Gait, M. J. ed., IRL Press, Oxford (1984)).

A variety of host-expression vector systems may be utilized to express the gene coding sequences of the invention. Such host-expression systems represent vehicles by which the coding sequences of interest may be produced and subsequently purified, but also represent cells that may, when transformed or transfected with the appropriate nucleotide coding sequences, exhibit the gene protein of the invention *in situ*. These include but are not limited to microorganisms such as bacteria (e.g., *E. coli*, *B. subtilis*) transformed with recombinant bacteriophage DNA, plasmid DNA or cosmid DNA expression vectors containing gene protein coding sequences; yeast (e.g. *Saccharomyces*, *Pichia*) transformed with recombinant yeast expression vectors containing the gene protein coding sequences; insect cell systems infected with recombinant virus expression vectors (e.g., baculovirus) containing the gene protein coding sequences; plant cell systems infected with recombinant virus expression vectors (e.g., cauliflower mosaic virus, CaMV; tobacco mosaic virus, TMV) or transformed with recombinant plasmid expression vectors (e.g., Ti plasmid) containing gene protein coding sequences; or mammalian cell systems (e.g. COS, CHO, BHK, 293, 3T3) harboring recombinant expression constructs containing promoters derived from the genome of mammalian cells (e.g., metallothionine promoter) or from mammalian viruses (e.g., the adenovirus late promoter; the vaccinia virus 7.5 K promoter).

In bacterial systems, a number of expression vectors may be advantageously selected depending upon the use intended for the gene protein being expressed. For example, when a large quantity of such a protein is to be produced, for the generation of antibodies or to screen peptide libraries, for example, vectors that direct the expression of high levels of fusion protein products that are readily purified may be desirable. Such vectors include, but are not limited, to the *E. coli* expression vector pUR278 (Ruther *et al.*, *EMBO J.*, 2:1791-94 (1983)), in which the gene protein coding sequence may be ligated individually into the vector in frame with the *lac Z* coding region so that a fusion protein is produced; pIN vectors (Inouye & Inouye, *Nucleic Acids Res.*, 13:3101-09 (1985); Van Heeke *et al.*, *J. Biol. Chem.*, 264:5503-9 (1989)); and the like. pGEX vectors may also be used to express foreign polypeptides as fusion proteins with glutathione S-transferase (GST). In general, such fusion proteins are soluble and can easily be purified from lysed cells by adsorption to glutathione-agarose beads followed by elution in the presence of free glutathione. The pGEX vectors are designed to include thrombin or factor Xa protease cleavage sites so that the cloned target gene protein can be released from the GST moiety.

In a preferred embodiment, full length cDNA sequences are appended with in-frame Bam HI sites at the amino terminus and Eco RI sites at the carboxyl terminus using standard PCR methodologies (Innis, *et al.* (eds) PCR Protocols: A Guide to Methods and Applications, Academic Press, San Diego (1990)) and ligated into the pGEX-2TK vector (Pharmacia, Uppsala, Sweden). The resulting cDNA construct contains a kinase recognition site at the amino terminus for radioactive

labeling and glutathione S-transferase sequences at the carboxyl terminus for affinity purification (Nilsson, *et al.*, *EMBO J.*, 4: 1075-80 (1985); Zabeau *et al.*, *EMBO J.*, 1: 1217-24 (1982)).

In an insect system, *Autographa californica* nuclear polyhedrosis virus (AcNPV) is used as a vector to express foreign genes. The virus grows in *Spodoptera frugiperda* cells. The gene coding
5 sequence may be cloned individually into non-essential regions (for example the polyhedrin gene) of the virus and placed under control of an AcNPV promoter (for example the polyhedrin promoter). Successful insertion of gene coding sequence will result in inactivation of the polyhedrin gene and production of non-occluded recombinant virus (*i.e.*, virus lacking the proteinaceous coat coded for by the polyhedrin gene). These recombinant viruses are then used to infect *Spodoptera frugiperda* cells
10 in which the inserted gene is expressed (*see, e.g.*, Smith, *et al.*, *J. Virol.* 46: 584-93 (1983); U.S. Patent No. 4,745,051).

In mammalian host cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, the gene coding sequence of interest may be ligated to an adenovirus transcription/translation control complex, *e.g.*, the late promoter and
15 tripartite leader sequence. This chimeric gene may then be inserted in the adenovirus genome by *in vitro* or *in vivo* recombination. Insertion in a non-essential region of the viral genome (*e.g.*, region E1 or E3) will result in a recombinant virus that is viable and capable of expressing gene protein in infected hosts. (*e.g.*, *see* Logan *et al.*, *Proc. Natl. Acad. Sci. USA*, 81:3655-59 (1984)). Specific initiation signals may also be required for efficient translation of inserted gene coding sequences.
20 These signals include the ATG initiation codon and adjacent sequences. In cases where an entire gene, including its own initiation codon and adjacent sequences, is inserted into the appropriate expression vector, no additional translational control signals may be needed. However, in cases where only a portion of the gene coding sequence is inserted, exogenous translational control signals, including, perhaps, the ATG initiation codon, must be provided. Furthermore, the initiation codon
25 must be in phase with the reading frame of the desired coding sequence to ensure translation of the entire insert. These exogenous translational control signals and initiation codons can be of a variety of origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of appropriate transcription enhancer elements, transcription terminators, etc. (*see* Bitter, *et al.*, *Methods in Enzymol.*, 153:516-44 (1987)).

In addition, a host cell strain may be chosen that modulates the expression of the inserted
30 sequences, or modifies and processes the gene product in the specific fashion desired. Such modifications (*e.g.*, glycosylation) and processing (*e.g.*, cleavage) of protein products may be important for the function of the protein. Different host cells have characteristic and specific mechanisms for the post-translational processing and modification of proteins. Appropriate cell lines
35 or host systems can be chosen to ensure the correct modification and processing of the foreign protein expressed. To this end, eukaryotic host cells that possess the cellular machinery for proper processing

of the primary transcript, glycosylation, and phosphorylation of the gene product may be used. Such mammalian host cells include but are not limited to CHO, VERO, BHK, HeLa, COS, MDCK, 293, 3T3, WI38, etc.

For long-term, high-yield production of recombinant proteins, stable expression is preferred. For example, cell lines that stably express the gene protein may be engineered. Rather than using expression vectors that contain viral origins of replication, host cells can be transformed with DNA controlled by appropriate expression control elements (*e.g.*, promoter, enhancer, sequences, transcription terminators, polyadenylation sites, etc.), and a selectable marker. Following the introduction of the foreign DNA, engineered cells may be allowed to grow for 1-2 days in an enriched media, and then are switched to a selective media. The selectable marker in the recombinant plasmid confers resistance to the selection and allows cells that stably integrate the plasmid into their chromosomes and grow, to form foci, which in turn can be cloned and expanded into cell lines. This method may advantageously be used to engineer cell lines that express the gene protein. Such engineered cell lines may be particularly useful in screening and evaluation of compounds that affect the endogenous activity of the gene protein.

In a preferred embodiment, timing and/or quantity of expression of the recombinant protein can be controlled using an inducible expression construct. Inducible constructs and systems for inducible expression of recombinant proteins will be well known to those skilled in the art. Examples of such inducible promoters or other gene regulatory elements include, but are not limited to, tetracycline, metallothionine, ecdysone, and other steroid-responsive promoters, rapamycin responsive promoters, and the like (No, *et al.*, *Proc. Natl. Acad. Sci. USA*, 93:3346-51 (1996); Furth, *et al.*, *Proc. Natl. Acad. Sci. USA*, 91:9302-6 (1994)). Additional control elements that can be used include promoters requiring specific transcription factors such as viral, particularly HIV, promoters. In one embodiment, a Tet inducible gene expression system is utilized. (Gossen *et al.*, *Proc. Natl. Acad. Sci. USA*, 89:5547-51 (1992); Gossen, *et al.*, *Science*, 268:1766-69 (1995)). Tet Expression Systems are based on two regulatory elements derived from the tetracycline-resistance operon of the *E. coli* Tn10 transposon—the tetracycline repressor protein (TetR) and the tetracycline operator sequence (*tetO*) to which TetR binds. Using such a system, expression of the recombinant protein is placed under the control of the *tetO* operator sequence and transfected or transformed into a host cell. In the presence of TetR, which is co-transfected into the host cell, expression of the recombinant protein is repressed due to binding of the TetR protein to the *tetO* regulatory element. High-level, regulated gene expression can then be induced in response to varying concentrations of tetracycline (Tc) or Tc derivatives such as doxycycline (Dox), which compete with *tetO* elements for binding to TetR. Constructs and materials for tet inducible gene expression are available commercially from CLONTECH Laboratories, Inc., Palo Alto, CA.

When used as a component in an assay system, the gene protein may be labeled, either directly or indirectly, to facilitate detection of a complex formed between the gene protein and a test substance. Any of a variety of suitable labeling systems may be used including but not limited to radioisotopes such as ^{125}I ; enzyme labeling systems that generate a detectable calorimetric signal or light when exposed to substrate; and fluorescent labels. Where recombinant DNA technology is used to produce the gene protein for such assay systems, it may be advantageous to engineer fusion proteins that can facilitate labeling, immobilization and/or detection.

Indirect labeling involves the use of a protein, such as a labeled antibody, which specifically binds to the gene product. Such antibodies include but are not limited to polyclonal, monoclonal, chimeric, single chain, Fab fragments and fragments produced by a Fab expression library.

Production of Antibodies

Described herein are methods for the production of antibodies capable of specifically recognizing one or more epitopes. Such antibodies may include, but are not limited to polyclonal antibodies, monoclonal antibodies (mAbs), humanized or chimeric antibodies, single chain antibodies, Fab fragments, F(ab)'_2 fragments, fragments produced by a Fab expression library, anti-idiotypic (anti-Id) antibodies, and epitope-binding fragments of any of the above. Such antibodies may be used, for example, in the detection of a target gene in a biological sample, or, alternatively, as a method for the inhibition of abnormal target gene activity. Thus, such antibodies may be utilized as part of disease treatment methods, and/or may be used as part of diagnostic techniques whereby patients may be tested for abnormal levels of target gene proteins, or for the presence of abnormal forms of such proteins.

For the production of antibodies, various host animals may be immunized by injection with the target gene, its expression product or a portion thereof. Such host animals may include but are not limited to rabbits, mice, rats, goats and chickens, to name but a few. Various adjuvants may be used to increase the immunological response, depending on the host species, including but not limited to Freund's (complete and incomplete), mineral gels such as aluminum hydroxide, surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, keyhole limpet hemocyanin, dinitrophenol, and potentially useful human adjuvants such as BCG (*bacille Calmette-Guerin*) and *Corynebacterium parvum*.

Polyclonal antibodies are heterogeneous populations of antibody molecules derived from the sera of animals immunized with an antigen, such as target gene product, or an antigenic functional derivative thereof. For the production of polyclonal antibodies, host animals such as those described above, may be immunized by injection with gene product supplemented with adjuvants as also described above.

Monoclonal antibodies, which are homogeneous populations of antibodies to a particular antigen, may be obtained by any technique that provides for the production of antibody molecules by

continuous cell lines in culture. These include, but are not limited to the hybridoma technique of Köhler and Milstein, *Nature*, 256:495-7 (1975); and U.S. Patent No. 4,376,110), the human B-cell hybridoma technique (Kosbor, *et al.*, *Immunology Today*, 4:72 (1983); Cote, *et al.*, *Proc. Natl. Acad. Sci. USA*, 80:2026-30 (1983)), and the EBV-hybridoma technique (Cole, *et al.*, in *Monoclonal Antibodies And Cancer Therapy*, Alan R. Liss, Inc., New York, pp. 77-96 (1985)). Such antibodies may be of any immunoglobulin class including IgG, IgM, IgE, IgA, IgD and any subclass thereof. The hybridoma producing the mAb of this invention may be cultivated *in vitro* or *in vivo*. Production of high titers of mAbs *in vivo* makes this the presently preferred method of production.

In addition, techniques developed for the production of "chimeric antibodies" (Morrison, *et al.*, *Proc. Natl. Acad. Sci.*, 81:6851-6855 (1984); Takeda, *et al.*, *Nature*, 314:452-54 (1985)) by splicing the genes from a mouse antibody molecule of appropriate antigen specificity together with genes from a human antibody molecule of appropriate biological activity can be used. A chimeric antibody is a molecule in which different portions are derived from different animal species, such as those having a variable region derived from a murine mAb and a human immunoglobulin constant region.

Alternatively, techniques described for the production of single chain antibodies (U.S. Patent No. 4,946,778; Bird, *Science* 242:423-26 (1988); Huston, *et al.*, *Proc. Natl. Acad. Sci. USA*, 85:5879-83 (1988); and Ward, *et al.*, *Nature*, 334:544-46 (1989)) can be adapted to produce gene-single chain antibodies. Single chain antibodies are typically formed by linking the heavy and light chain fragments of the Fv region via an amino acid bridge, resulting in a single chain polypeptide.

Antibody fragments that recognize specific epitopes may be generated by known techniques. For example, such fragments include but are not limited to: the F(ab')₂ fragments that can be produced by pepsin digestion of the antibody molecule and the Fab fragments that can be generated by reducing the disulfide bridges of the F(ab')₂ fragments. Alternatively, Fab expression libraries may be constructed (Huse, *et al.*, *Science*, 246:1275-81 (1989)) to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity.

Screening Methods

The present invention may be employed in a process for screening for agents such as agonists, *i.e.* agents that bind to and activate target polypeptides, or antagonists, *i.e.* inhibit the activity or interaction of target polypeptides with its ligand. Thus, polypeptides of the invention may also be used to assess the binding of small molecule substrates and ligands in, for example, cells, cell-free preparations, chemical libraries, and natural product mixtures as known in the art. Any methods routinely used to identify and screen for agents that can modulate receptors may be used in accordance with the present invention.

The present invention provides methods for identifying and screening for agents that modulate target gene expression or function. More particularly, cells that contain and express target

gene sequences may be used to screen for therapeutic agents. Such cells may include non-recombinant monocyte cell lines, such as U937 (ATCC# CRL-1593), THP-1 (ATCC# TIB-202), and P388D1 (ATCC# TIB-63); endothelial cells such as HUVEC's and bovine aortic endothelial cells (BAEC's); as well as generic mammalian cell lines such as HeLa cells and COS cells, *e.g.*, COS-7
5 (ATCC# CRL-1651). Further, such cells may include recombinant, transgenic cell lines. For example, the transgenic mice of the invention may be used to generate cell lines, containing one or more cell types involved in a disease, that can be used as cell culture models for that disorder. While cells, tissues, and primary cultures derived from the disease transgenic animals of the invention may be utilized, the generation of continuous cell lines is preferred. For examples of techniques that may
10 be used to derive a continuous cell line from the transgenic animals, see Small, *et al.*, *Mol. Cell Biol.*, 5:642-48 (1985).

Target gene sequences may be introduced into, and overexpressed in, the genome of the cell of interest. In order to overexpress a target gene sequence, the coding portion of the target gene sequence may be ligated to a regulatory sequence that is capable of driving gene expression in the cell
15 type of interest. Such regulatory regions will be well known to those of skill in the art, and may be utilized in the absence of undue experimentation. Target gene sequences may also be disrupted or underexpressed. Cells having targeted gene disruptions or underexpressed target gene sequences may be used, for example, to screen for agents capable of affecting alternative pathways that compensate for any loss of function attributable to the disruption or underexpression.

In vitro systems may be designed to identify compounds capable of binding the target gene products. Such compounds may include, but are not limited to, peptides made of D-and/or L-configuration amino acids (in, for example, the form of random peptide libraries; (*see e.g.*, Lam, *et al.*, *Nature*, 354:82-4 (1991)), phosphopeptides (in, for example, the form of random or partially degenerate, directed phosphopeptide libraries; *see, e.g.*, Songyang, *et al.*, *Cell*, 72:767-78 (1993)),
25 antibodies, and small organic or inorganic molecules. Compounds identified may be useful, for example, in modulating the activity of target gene proteins, preferably mutant target gene proteins; elaborating the biological function of the target gene protein; or screening for compounds that disrupt normal target gene interactions or themselves disrupt such interactions.

The principle of the assays used to identify compounds that bind to the target gene protein
30 involves preparing a reaction mixture of the target gene protein and the test compound under conditions and for a time sufficient to allow the two components to interact and bind, thus forming a complex that can be removed and/or detected in the reaction mixture. These assays can be conducted in a variety of ways. For example, one method to conduct such an assay would involve anchoring the target gene protein or the test substance onto a solid phase and detecting target protein/test substance
35 complexes anchored on the solid phase at the end of the reaction. In one embodiment of such a

method, the target gene protein may be anchored onto a solid surface, and the test compound, which is not anchored, may be labeled, either directly or indirectly.

In practice, microtitre plates are conveniently utilized. The anchored component may be immobilized by non-covalent or covalent attachments. Non-covalent attachment may be accomplished simply by coating the solid surface with a solution of the protein and drying.
5 Alternatively, an immobilized antibody, preferably a monoclonal antibody, specific for the protein may be used to anchor the protein to the solid surface. The surfaces may be prepared in advance and stored.

In order to conduct the assay, the nonimmobilized component is added to the coated surface containing the anchored component. After the reaction is complete, unreacted components are removed (*e.g.*, by washing) under conditions such that any complexes formed will remain immobilized on the solid surface. The detection of complexes anchored on the solid surface can be accomplished in a number of ways. Where the previously nonimmobilized component is pre-labeled, the detection of label immobilized on the surface indicates that complexes were formed. Where the
10 previously nonimmobilized component is not pre-labeled, an indirect label can be used to detect complexes anchored on the surface; *e.g.*, using a labeled antibody specific for the previously nonimmobilized component (the antibody, in turn, may be directly labeled or indirectly labeled with a
15 labeled anti-Ig antibody).

Alternatively, a reaction can be conducted in a liquid phase, the reaction products separated from unreacted components, and complexes detected; *e.g.*, using an immobilized antibody specific for
20 target gene product or the test compound to anchor any complexes formed in solution, and a labeled antibody specific for the other component of the possible complex to detect anchored complexes.

Compounds that are shown to bind to a particular target gene product through one of the methods described above can be further tested for their ability to elicit a biochemical response from the target gene protein. Agonists, antagonists and/or inhibitors of the expression product can be
25 identified utilizing assays well known in the art.

Antisense, Ribozymes, and Antibodies

Other agents that may be used as therapeutics include the target gene, its expression product(s) and functional fragments thereof. Additionally, agents that reduce or inhibit mutant target
30 gene activity may be used to ameliorate disease symptoms. Such agents include antisense, ribozyme, and triple helix molecules. Techniques for the production and use of such molecules are well known to those of skill in the art.

Anti-sense RNA and DNA molecules act to directly block the translation of mRNA by hybridizing to targeted mRNA and preventing protein translation. With respect to antisense DNA,
35 oligodeoxyribonucleotides derived from the translation initiation site, *e.g.*, between the -10 and +10 regions of the target gene nucleotide sequence of interest, are preferred.

Ribozymes are enzymatic RNA molecules capable of catalyzing the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by an endonucleolytic cleavage. The composition of ribozyme molecules must include one or more sequences complementary to the target gene mRNA, and must include the well known catalytic sequence responsible for mRNA cleavage. For this sequence, see U.S. Patent No. 5,093,246, which is incorporated by reference herein in its entirety. As such within the scope of the invention are engineered hammerhead motif ribozyme molecules that specifically and efficiently catalyze endonucleolytic cleavage of RNA sequences encoding target gene proteins.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the molecule of interest for ribozyme cleavage sites that include the following sequences, GUA, GUU and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides corresponding to the region of the target gene containing the cleavage site may be evaluated for predicted structural features, such as secondary structure, that may render the oligonucleotide sequence unsuitable. The suitability of candidate sequences may also be evaluated by testing their accessibility to hybridization with complementary oligonucleotides, using ribonuclease protection assays.

Nucleic acid molecules to be used in triple helix formation for the inhibition of transcription should be single stranded and composed of deoxyribonucleotides. The base composition of these oligonucleotides must be designed to promote triple helix formation via Hoogsteen base pairing rules, which generally require sizeable stretches of either purines or pyrimidines to be present on one strand of a duplex. Nucleotide sequences may be pyrimidine-based, which will result in TAT and CGC triplets across the three associated strands of the resulting triple helix. The pyrimidine-rich molecules provide base complementarity to a purine-rich region of a single strand of the duplex in a parallel orientation to that strand. In addition, nucleic acid molecules may be chosen that are purine-rich, for example, containing a stretch of G residues. These molecules will form a triple helix with a DNA duplex that is rich in GC pairs, in which the majority of the purine residues are located on a single strand of the targeted duplex, resulting in GGC triplets across the three strands in the triplex.

Alternatively, the potential sequences that can be targeted for triple helix formation may be increased by creating a so called "switchback" nucleic acid molecule. Switchback molecules are synthesized in an alternating 5'-3', 3'-5' manner, such that they base pair with first one strand of a duplex and then the other, eliminating the necessity for a sizeable stretch of either purines or pyrimidines to be present on one strand of a duplex.

It is possible that the antisense, ribozyme, and/or triple helix molecules described herein may reduce or inhibit the transcription (triple helix) and/or translation (antisense, ribozyme) of mRNA produced by both normal and mutant target gene alleles. In order to ensure that substantially normal

levels of target gene activity are maintained, nucleic acid molecules that encode and express target gene polypeptides exhibiting normal activity may be introduced into cells that do not contain sequences susceptible to whatever antisense, ribozyme, or triple helix treatments are being utilized. Alternatively, it may be preferable to coadminister normal target gene protein into the cell or tissue in order to maintain the requisite level of cellular or tissue target gene activity.

Anti-sense RNA and DNA, ribozyme, and triple helix molecules of the invention may be prepared by any method known in the art for the synthesis of DNA and RNA molecules. These include techniques for chemically synthesizing oligodeoxyribonucleotides and oligoribonucleotides well known in the art such as for example solid phase phosphoramidite chemical synthesis.

Alternatively, RNA molecules may be generated by *in vitro* and *in vivo* transcription of DNA sequences encoding the antisense RNA molecule. Such DNA sequences may be incorporated into a wide variety of vectors that incorporate suitable RNA polymerase promoters such as the T7 or SP6 polymerase promoters. Alternatively, antisense cDNA constructs that synthesize antisense RNA constitutively or inducibly, depending on the promoter used, can be introduced stably into cell lines.

Various well-known modifications to the DNA molecules may be introduced as a means of increasing intracellular stability and half-life. Possible modifications include but are not limited to the addition of flanking sequences of ribonucleotides or deoxyribonucleotides to the 5' and/or 3' ends of the molecule or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the oligodeoxyribonucleotide backbone.

Antibodies that are both specific for target gene protein, and in particular, mutant gene protein, and interfere with its activity may be used to inhibit mutant target gene function. Such antibodies may be generated against the proteins themselves or against peptides corresponding to portions of the proteins using standard techniques known in the art and as also described herein. Such antibodies include but are not limited to polyclonal, monoclonal, Fab fragments, single chain antibodies, chimeric antibodies, etc.

In instances where the target gene protein is intracellular and whole antibodies are used, internalizing antibodies may be preferred. However, lipofectin liposomes may be used to deliver the antibody or a fragment of the Fab region that binds to the target gene epitope into cells. Where fragments of the antibody are used, the smallest inhibitory fragment that binds to the target or expanded target protein's binding domain is preferred. For example, peptides having an amino acid sequence corresponding to the domain of the variable region of the antibody that binds to the target gene protein may be used. Such peptides may be synthesized chemically or produced via recombinant DNA technology using methods well known in the art (*see, e.g.,* Creighton, *Proteins: Structures and Molecular Principles* (1984) W.H. Freeman, New York 1983, *supra*; and Sambrook, *et al.*, 1989, *supra*). Alternatively, single chain neutralizing antibodies that bind to intracellular target gene epitopes may also be administered. Such single chain antibodies may be administered, for

example, by expressing nucleotide sequences encoding single-chain antibodies within the target cell population by utilizing, for example, techniques such as those described in Marasco, *et al.*, *Proc. Natl. Acad. Sci. USA*, 90:7889-93 (1993).

RNA sequences encoding target gene protein may be directly administered to a patient exhibiting disease symptoms, at a concentration sufficient to produce a level of target gene protein such that disease symptoms are ameliorated. Patients may be treated by gene replacement therapy. One or more copies of a normal target gene, or a portion of the gene that directs the production of a normal target gene protein with target gene function, may be inserted into cells using vectors that include, but are not limited to adenovirus, adeno-associated virus, and retrovirus vectors, in addition to other particles that introduce DNA into cells, such as liposomes. Additionally, techniques such as those described above may be utilized for the introduction of normal target gene sequences into human cells.

Cells, preferably, autologous cells, containing normal target gene expressing gene sequences may then be introduced or reintroduced into the patient at positions that allow for the amelioration of disease symptoms.

Pharmaceutical Compositions, Effective Dosages, and Routes of Administration

The identified compounds that inhibit target mutant gene expression, synthesis and/or activity can be administered to a patient at therapeutically effective doses to treat or ameliorate the disease. A therapeutically effective dose refers to that amount of the compound sufficient to result in amelioration of symptoms of the disease.

Toxicity and therapeutic efficacy of such compounds can be determined by standard pharmaceutical procedures in cell cultures or experimental animals, *e.g.*, for determining the LD₅₀ (the dose lethal to 50% of the population) and the ED₅₀ (the dose therapeutically effective in 50% of the population). The dose ratio between toxic and therapeutic effects is the therapeutic index and it can be expressed as the ratio LD₅₀/ED₅₀. Compounds that exhibit large therapeutic indices are preferred. While compounds that exhibit toxic side effects may be used, care should be taken to design a delivery system that targets such compounds to the site of affected tissue in order to minimize potential damage to uninfected cells and, thereby, reduce side effects.

The data obtained from the cell culture assays and animal studies can be used in formulating a range of dosage for use in humans. The dosage of such compounds lies preferably within a range of circulating concentrations that include the ED₅₀ with little or no toxicity. The dosage may vary within this range depending upon the dosage form employed and the route of administration utilized. For any compound used in the method of the invention, the therapeutically effective dose can be estimated initially from cell culture assays. A dose may be formulated in animal models to achieve a circulating plasma concentration range that includes the IC₅₀ (*i.e.*, the concentration of the test compound that achieves a half-maximal inhibition of symptoms) as determined in cell culture. Such information can

be used to more accurately determine useful doses in humans. Levels in plasma may be measured, for example, by high performance liquid chromatography.

Pharmaceutical compositions for use in accordance with the present invention may be formulated in conventional manner using one or more physiologically acceptable carriers or excipients. Thus, the compounds and their physiologically acceptable salts and solvates may be formulated for administration by inhalation or insufflation (either through the mouth or the nose) or oral, buccal, parenteral, topical, subcutaneous, intraperitoneal, intravenous, intrapleural, intraocular, intraarterial, or rectal administration. It is also contemplated that pharmaceutical compositions may be administered with other products that potentiate the activity of the compound and optionally, may include other therapeutic ingredients.

For oral administration, the pharmaceutical compositions may take the form of, for example, tablets or capsules prepared by conventional means with pharmaceutically acceptable excipients such as binding agents (*e.g.*, pregelatinized maize starch, polyvinylpyrrolidone or hydroxypropyl methylcellulose); fillers (*e.g.*, lactose, microcrystalline cellulose or calcium hydrogen phosphate); lubricants (*e.g.*, magnesium stearate, talc or silica); disintegrants (*e.g.*, potato starch or sodium starch glycolate); or wetting agents (*e.g.*, sodium lauryl sulphate). The tablets may be coated by methods well known in the art. Liquid preparations for oral administration may take the form of, for example, solutions, syrups or suspensions, or they may be presented as a dry product for constitution with water or other suitable vehicle before use. Such liquid preparations may be prepared by conventional means with pharmaceutically acceptable additives such as suspending agents (*e.g.*, sorbitol syrup, cellulose derivatives or hydrogenated edible fats); emulsifying agents (*e.g.*, lecithin or acacia); non-aqueous vehicles (*e.g.*, almond oil, oily esters, ethyl alcohol or fractionated vegetable oils); and preservatives (*e.g.*, methyl or propyl-p-hydroxybenzoates or sorbic acid). The preparations may also contain buffer salts, flavoring, coloring and sweetening agents as appropriate.

Preparations for oral administration may be suitably formulated to give controlled release of the active compound.

For buccal administration the compositions may take the form of tablets or lozenges formulated in conventional manner.

For administration by inhalation, the compounds for use according to the present invention are conveniently delivered in the form of an aerosol spray presentation from pressurized packs or a nebulizer, with the use of a suitable propellant, *e.g.*, dichlorodifluoromethane, trichlorofluoromethane, dichlorotetrafluoroethane, carbon dioxide or other suitable gas. In the case of a pressurized aerosol the dosage unit may be determined by providing a valve to deliver a metered amount. Capsules and cartridges of *e.g.* gelatin for use in an inhaler or insufflator may be formulated containing a powder mix of the compound and a suitable powder base such as lactose or starch.

The compounds may be formulated for parenteral administration by injection, *e.g.*, by bolus injection or continuous infusion. Formulations for injection may be presented in unit dosage form, *e.g.*, in ampoules or in multi-dose containers, with an added preservative. The compositions may take such forms as suspensions, solutions or emulsions in oily or aqueous vehicles, and may contain
5 formulatory agents such as suspending, stabilizing and/or dispersing agents. Alternatively, the active ingredient may be in powder form for constitution with a suitable vehicle, *e.g.*, sterile pyrogen-free water, before use.

The compounds may also be formulated in rectal compositions such as suppositories or retention enemas, *e.g.*, containing conventional suppository bases such as cocoa butter or other
10 glycerides. Oral ingestion is possibly the easiest method of taking any medication. Such a route of administration, is generally simple and straightforward and is frequently the least inconvenient or unpleasant route of administration from the patient's point of view. However, this involves passing the material through the stomach, which is a hostile environment for many materials, including proteins and other biologically active compositions. As the acidic, hydrolytic and proteolytic
15 environment of the stomach has evolved efficiently to digest proteinaceous materials into amino acids and oligopeptides for subsequent anabolism, it is hardly surprising that very little or any of a wide variety of biologically active proteinaceous material, if simply taken orally, would survive its passage through the stomach to be taken up by the body in the small intestine. The result, is that many proteinaceous medicaments must be taken in through another method, such as parenterally, often by
20 subcutaneous, intramuscular or intravenous injection.

Pharmaceutical compositions may also include various buffers (*e.g.*, Tris, acetate, phosphate), solubilizers (*e.g.*, Tween, Polysorbate), carriers such as human serum albumin, preservatives (thimerosal, benzyl alcohol) and anti-oxidants such as ascorbic acid in order to stabilize
pharmaceutical activity. The stabilizing agent may be a detergent, such as tween-20, tween-80, NP-
25 40 or Triton X-100. EBP may also be incorporated into particulate preparations of polymeric compounds for controlled delivery to a patient over an extended period of time. A more extensive survey of components in pharmaceutical compositions is found in Remington's Pharmaceutical Sciences, 18th ed., A. R. Gennaro, ed., Mack Publishing, Easton, Pa. (1990).

In addition to the formulations described previously, the compounds may also be formulated
30 as a depot preparation. Such long acting formulations may be administered by implantation (for example subcutaneously or intramuscularly) or by intramuscular injection. Thus, for example, the compounds may be formulated with suitable polymeric or hydrophobic materials (for example as an emulsion in an acceptable oil) or ion exchange resins, or as sparingly soluble derivatives, for example, as a sparingly soluble salt.

35 The compositions may, if desired, be presented in a pack or dispenser device that may contain one or more unit dosage forms containing the active ingredient. The pack may for example comprise

metal or plastic foil, such as a blister pack. The pack or dispenser device may be accompanied by instructions for administration.

Diagnostics

5 A variety of methods may be employed to diagnose disease conditions associated with the target gene. Specifically, reagents may be used, for example, for the detection of the presence of target gene mutations, or the detection of either over or under expression of target gene mRNA.

10 According to the diagnostic and prognostic method of the present invention, alteration of the wild-type target gene locus is detected. In addition, the method can be performed by detecting the wild-type target gene locus and confirming the lack of a predisposition or neoplasia. "Alteration of a wild-type gene" encompasses all forms of mutations including deletions, insertions and point mutations in the coding and noncoding regions. Deletions may be of the entire gene or only a portion of the gene. Point mutations may result in stop codons, frameshift mutations or amino acid substitutions. Somatic mutations are those that occur only in certain tissues, *e.g.*, in tumor tissue, and are not inherited in the germline. Germline mutations can be found in any of a body's tissues and are inherited. If only a single allele is somatically mutated, an early neoplastic state may be indicated. However, if both alleles are mutated, then a late neoplastic state may be indicated. The finding of gene mutations thus provides both diagnostic and prognostic information. A target gene allele that is not deleted (*e.g.*, that found on the sister chromosome to a chromosome carrying a target gene deletion) can be screened for other mutations, such as insertions, small deletions, and point mutations. Mutations found in tumor tissues may be linked to decreased expression of the target gene product. However, mutations leading to non-functional gene products may also be linked to a cancerous state. Point mutational events may occur in regulatory regions, such as in the promoter of the gene, leading to loss or diminution of expression of the mRNA. Point mutations may also abolish proper RNA processing, leading to loss of expression of the target gene product, or a decrease in mRNA stability or translation efficiency.

25 One test available for detecting mutations in a candidate locus is to directly compare genomic target sequences from cancer patients with those from a control population. Alternatively, one could sequence messenger RNA after amplification, *e.g.*, by PCR, thereby eliminating the necessity of determining the exon structure of the candidate gene. Mutations from cancer patients falling outside the coding region of the target gene can be detected by examining the non-coding regions, such as introns and regulatory sequences near or within the target gene. An early indication that mutations in noncoding regions are important may come from Northern blot experiments that reveal messenger RNA molecules of abnormal size or abundance in cancer patients as compared to control individuals.

35 The methods described herein may be performed, for example, by utilizing pre-packaged diagnostic kits comprising at least one specific gene nucleic acid or anti-gene antibody reagent

described herein, which may be conveniently used, *e.g.*, in clinical settings, to diagnose patients exhibiting disease symptoms or at risk for developing disease.

Any cell type or tissue in which the gene is expressed may be utilized in the diagnostics described below. DNA or RNA from the cell type or tissue to be analyzed may easily be isolated
5 using procedures that are well known to those in the art. Diagnostic procedures may also be performed *in situ* directly upon tissue sections (fixed and/or frozen) of patient tissue obtained from biopsies or resections, such that no nucleic acid purification is necessary. Nucleic acid reagents may be used as probes and/or primers for such *in situ* procedures (*see*, for example, Nuovo, *PCR In Situ Hybridization: Protocols and Applications*, Raven Press, N.Y. (1992)).

10 Gene nucleotide sequences, either RNA or DNA, may, for example, be used in hybridization or amplification assays of biological samples to detect disease-related gene structures and expression. Such assays may include, but are not limited to, Southern or Northern analyses, restriction fragment length polymorphism assays, single stranded conformational polymorphism analyses, *in situ* hybridization assays, and polymerase chain reaction analyses. Such analyses may reveal both quantitative
15 aspects of the expression pattern of the gene, and qualitative aspects of the gene expression and/or gene composition. That is, such aspects may include, for example, point mutations, insertions, deletions, chromosomal rearrangements, and/or activation or inactivation of gene expression.

Preferred diagnostic methods for the detection of gene-specific nucleic acid molecules may involve for example, contacting and incubating nucleic acids, derived from the cell type or tissue
20 being analyzed, with one or more labeled nucleic acid reagents under conditions favorable for the specific annealing of these reagents to their complementary sequences within the nucleic acid molecule of interest. Preferably, the lengths of these nucleic acid reagents are at least 9 to 30 nucleotides. After incubation, all non-annealed nucleic acids are removed from the nucleic acid:fingerprint molecule hybrid. The presence of nucleic acids from the fingerprint tissue that have
25 hybridized, if any such molecules exist, is then detected. Using such a detection scheme, the nucleic acid from the tissue or cell type of interest may be immobilized, for example, to a solid support such as a membrane, or a plastic surface such as that on a microtitre plate or polystyrene beads. In this case, after incubation, non-annealed, labeled nucleic acid reagents are easily removed. Detection of the remaining, annealed, labeled nucleic acid reagents is accomplished using standard techniques
30 well-known to those in the art.

Alternative diagnostic methods for the detection of gene-specific nucleic acid molecules may involve their amplification, *e.g.*, by PCR (the experimental embodiment set forth in Mullis U.S. Patent No. 4,683,202 (1987)), ligase chain reaction (Barany, *Proc. Natl. Acad. Sci. USA*, 88:189-93 (1991)), self sustained sequence replication (Guatelli, *et al.*, *Proc. Natl. Acad. Sci. USA*, 87:1874-78 (1990)),
35 transcriptional amplification system (Kwoh, *et al.*, *Proc. Natl. Acad. Sci. USA*, 86:1173-77 (1989)), Q-Beta Replicase (Lizardi *et al.*, *Bio/Technology*, 6:1197 (1988)), or any other nucleic acid

amplification method, followed by the detection of the amplified molecules using techniques well known to those of skill in the art. These detection schemes are especially useful for the detection of nucleic acid molecules if such molecules are present in very low numbers.

In one embodiment of such a detection scheme, a cDNA molecule is obtained from an RNA molecule of interest (*e.g.*, by reverse transcription of the RNA molecule into cDNA). Cell types or tissues from which such RNA may be isolated include any tissue in which wild type fingerprint gene is known to be expressed. A sequence within the cDNA is then used as the template for a nucleic acid amplification reaction, such as a PCR amplification reaction, or the like. The nucleic acid reagents used as synthesis initiation reagents (*e.g.*, primers) in the reverse transcription and nucleic acid amplification steps of this method may be chosen from among the gene nucleic acid reagents described herein. The preferred lengths of such nucleic acid reagents are at least 15-30 nucleotides. For detection of the amplified product, the nucleic acid amplification may be performed using radioactively or non-radioactively labeled nucleotides. Alternatively, enough amplified product may be made such that the product may be visualized by standard ethidium bromide staining or by utilizing any other suitable nucleic acid staining method.

Antibodies directed against wild type or mutant gene peptides may also be used as disease diagnostics and prognostics. Such diagnostic methods, may be used to detect abnormalities in the level of gene protein expression, or abnormalities in the structure and/or tissue, cellular, or subcellular location of fingerprint gene protein. Structural differences may include, for example, differences in the size, electronegativity, or antigenicity of the mutant fingerprint gene protein relative to the normal fingerprint gene protein.

Protein from the tissue or cell type to be analyzed may easily be detected or isolated using techniques that are well known to those of skill in the art, including but not limited to western blot analysis. For a detailed explanation of methods for carrying out western blot analysis, see Sambrook, *et al.* (1989) *supra*, at Chapter 18. The protein detection and isolation methods employed herein may also be such as those described in Harlow and Lane, for example, (*Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York (1988)).

Preferred diagnostic methods for the detection of wild type or mutant gene peptide molecules may involve, for example, immunoassays wherein fingerprint gene peptides are detected by their interaction with an anti-fingerprint gene-specific peptide antibody.

For example, antibodies, or fragments of antibodies useful in the present invention may be used to quantitatively or qualitatively detect the presence of wild type or mutant gene peptides. This can be accomplished, for example, by immunofluorescence techniques employing a fluorescently labeled antibody (see below) coupled with light microscopic, flow cytometric, or fluorimetric detection. Such techniques are especially preferred if the fingerprint gene peptides are expressed on the cell surface.

The antibodies (or fragments thereof) useful in the present invention may, additionally, be employed histologically, as in immunofluorescence or immunoelectron microscopy, for *in situ* detection of fingerprint gene peptides. *In situ* detection may be accomplished by removing a histological specimen from a patient, and applying thereto a labeled antibody of the present invention. The antibody (or fragment) is preferably applied by overlaying the labeled antibody (or fragment) onto a biological sample. Through the use of such a procedure, it is possible to determine not only the presence of the fingerprint gene peptides, but also their distribution in the examined tissue. Using the present invention, those of ordinary skill will readily perceive that any of a wide variety of histological methods (such as staining procedures) can be modified in order to achieve such *in situ* detection.

Immunoassays for wild type, mutant, or expanded fingerprint gene peptides typically comprise incubating a biological sample, such as a biological fluid, a tissue extract, freshly harvested cells, or cells that have been incubated in tissue culture, in the presence of a detectably labeled antibody capable of identifying fingerprint gene peptides, and detecting the bound antibody by any of a number of techniques well known in the art.

The biological sample may be brought in contact with and immobilized onto a solid phase support or carrier such as nitrocellulose, or other solid support that is capable of immobilizing cells, cell particles or soluble proteins. The support may then be washed with suitable buffers followed by treatment with the detectably labeled gene-specific antibody. The solid phase support may then be washed with the buffer a second time to remove unbound antibody. The amount of bound label on solid support may then be detected by conventional means.

The terms "solid phase support or carrier" are intended to encompass any support capable of binding an antigen or an antibody. Well-known supports or carriers include glass, polystyrene, polypropylene, polyethylene, dextran, nylon, amylases, natural and modified celluloses, polyacrylamides, gabbros, and magnetite. The nature of the carrier can be either soluble to some extent or insoluble for the purposes of the present invention. The support material may have virtually any possible structural configuration so long as the coupled molecule is capable of binding to an antigen or antibody. Thus, the support configuration may be spherical, as in a bead, or cylindrical, as in the inside surface of a test tube, or the external surface of a rod. Alternatively, the surface may be flat such as a sheet, test strip, etc. Preferred supports include polystyrene beads. Those skilled in the art will know many other suitable carriers for binding antibody or antigen, or will be able to ascertain the same by use of routine experimentation.

The binding activity of a given lot of anti-wild type or -mutant fingerprint gene peptide antibody may be determined according to well known methods. Those skilled in the art will be able to determine operative and optimal assay conditions for each determination by employing routine experimentation.

One of the ways in which the gene peptide-specific antibody can be detectably labeled is by linking the same to an enzyme and using it in an enzyme immunoassay (EIA) (Voller, *Ric Clin Lab*, 8:289-98 (1978) ["The Enzyme Linked Immunosorbent Assay (ELISA)", *Diagnostic Horizons* 2:1-7, 1978, Microbiological Associates Quarterly Publication, Walkersville, Md.]; Voller, *et al.*, *J. Clin. Pathol.*, 31:507-20 (1978); Butler, *Meth. Enzymol.*, 73:482-523 (1981); Maggio (ed.), *Enzyme Immunoassay*, CRC Press, Boca Raton, Fla. (1980); Ishikawa, *et al.*, (eds.) *Enzyme Immunoassay*, Igaku-Shoin, Tokyo (1981)). The enzyme that is bound to the antibody will react with an appropriate substrate, preferably a chromogenic substrate, in such a manner as to produce a chemical moiety that can be detected, for example, by spectrophotometric, fluorimetric or by visual means. Enzymes that can be used to detectably label the antibody include, but are not limited to, malate dehydrogenase, staphylococcal nuclease, delta-5-steroid isomerase, yeast alcohol dehydrogenase, alpha-glycerophosphate dehydrogenase, triose phosphate isomerase, horseradish peroxidase, alkaline phosphatase, asparaginase, glucose oxidase, beta-galactosidase, ribonuclease, urease, catalase, glucose-6-phosphate dehydrogenase, glucoamylase and acetylcholinesterase. The detection can be accomplished by colorimetric methods that employ a chromogenic substrate for the enzyme. Detection may also be accomplished by visual comparison of the extent of enzymatic reaction of a substrate in comparison with similarly prepared standards.

Detection may also be accomplished using any of a variety of other immunoassays. For example, by radioactively labeling the antibodies or antibody fragments, it is possible to detect fingerprint gene wild type, mutant, or expanded peptides through the use of a radioimmunoassay (RIA) (*see, e.g.*, Weintraub, B., *Principles of Radioimmunoassays*, Seventh Training Course on Radioligand Assay Techniques, The Endocrine Society, March, 1986). The radioactive isotope can be detected by such means as the use of a gamma counter or a scintillation counter or by autoradiography.

It is also possible to label the antibody with a fluorescent compound. When the fluorescently labeled antibody is exposed to light of the proper wave length, its presence can then be detected due to fluorescence. Among the most commonly used fluorescent labeling compounds are fluorescein isothiocyanate, rhodamine, phycoerythrin, phycocyanin, allophycocyanin, o-phthaldehyde and fluorescamine.

The antibody can also be detectably labeled using fluorescence emitting metals such as ^{152}Eu , or others of the lanthanide series. These metals can be attached to the antibody using such metal chelating groups as diethylenetriaminepentaacetic acid (DTPA) or ethylenediamine-tetraacetic acid (EDTA).

The antibody also can be detectably labeled by coupling it to a chemiluminescent compound. The presence of the chemiluminescent-tagged antibody is then determined by detecting the presence of luminescence that arises during the course of a chemical reaction. Examples of particularly useful

chemiluminescent labeling compounds are luminol, isoluminol, therromatic acridinium ester, imidazole, acridinium salt and oxalate ester.

Likewise, a bioluminescent compound may be used to label the antibody of the present invention. Bioluminescence is a type of chemiluminescence found in biological systems in which a catalytic protein increases the efficiency of the chemiluminescent reaction. The presence of a bioluminescent protein is determined by detecting the presence of luminescence. Important bioluminescent compounds for purposes of labeling are luciferin, luciferase and aequorin.

Throughout this application, various publications, patents and published patent applications are referred to by an identifying citation. The disclosures of these publications, patents and published patent specifications referenced in this application are hereby incorporated by reference into the present disclosure to more fully describe the state of the art to which this invention pertains.

The following examples are intended only to illustrate the present invention and should in no way be construed as limiting the subject invention.

Examples

15 **Example 1: Generation and Analysis of Mice Comprising CHD2-like Gene Disruptions**

Targeting Construct. To investigate the role of CHD2-like molecules, disruptions in CHD2-like genes were produced by homologous recombination. More particularly, as shown in Figure 2, a CHD2-like-specific targeting construct having the ability to disrupt or modify CHD2-like genes, specifically comprising SEQ ID NO:1 was created using as the targeting arms (homologous sequences) in the construct, the oligonucleotide sequences identified herein as SEQ ID NO:2 or SEQ ID NO:3.

Transgenic Mice. ES cells derived from the 129/SvJ x 129/Sv-CP mouse substrain were used to generate chimeric mice. F1 mice were generated by breeding with C57BL/6 females. The resultant F1N0 heterozygotes were backcrossed to C57BL/6 mice to generate F1N1 heterozygotes. F2N1 homozygous mutant mice were produced by intercrossing F1N1 heterozygous males and females.

Phenotypic Analysis. The transgenic mice were analyzed for phenotypic changes and expression patterns. The homozygous mice demonstrated at least one of the following phenotypes: *Expression.*

Lac-Z. Tissues of the transgenic animals were analyzed for expression of the target gene. Organs from one heterozygous male and one heterozygous female were frozen, sectioned (10 μ m), stained and analyzed for lacZ expression using X-Gal as a substrate for beta-galactosidase. Nuclear Fast Red was used for counterstaining.

Organs and tissues collected and frozen: brain, sciatic nerve, eye, Harderian glands, thymus, spleen, lymph nodes, bone marrow, aorta, heart, lung, liver, gallbladder, pancreas, kidney, urinary bladder, trachea, larynx, esophagus, thyroid gland, pituitary gland, adrenal glands, salivary glands, stomach, small and large intestines, tongue, skeletal muscle, skin and reproductive system.

In addition, the brain of the heterozygous female was analyzed for lacZ expression as wholemount. The dissected brain was cut longitudinally, fixed and stained using X-Gal as a substrate for beta-galactosidase. To stop the reaction the brain was washed in PBS and fixed in PBS-buffered formaldehyde.

Wild type control tissues were stained for X-gal to reveal background or signals due to endogenous beta-galactosidase activity. The following tissues show staining in the wild-type control sections and are therefore not suitable for X-gal staining: small and large intestines, stomach, vas deferens and epididymis. It has been previously reported that these organs contain high levels of endogenous beta-galactosidase activity.

The results were as follows: X-Gal staining was detected only in salivary and coagulating glands. Comparing lacZ expression with data from RT-PCR analysis (below) suggests that the lacZ reporter gene does not fully reflect the endogenous expression pattern. Specific details of the staining pattern observed in each tissue are given below.

Salivary Glands. Many epithelial cells of the salivary glands displayed X-Gal staining.

Male Reproductive System: Coagulating Glands. A few epithelial cells showed moderate X-Gal staining.

5 *RT-PCR.* Total RNA was isolated from the organs or tissues from adult C57Bl/6 wild type mice. RNA was DNaseI treated, and reverse transcribed using random primers. The resulting cDNA was checked for the absence of genomic contamination using primers specific to non-transcribed genomic mouse DNA. cDNAs were balanced for concentration using HPRT primers. RNA transcripts were detected in eye, lung, pancreas, kidneys, lymph nodes, skin, pituitary gland, adrenal gland, salivary gland, skeletal muscle, tongue, stomach, small intestine, large intestine, cecum, coagulating
10 gland and prostate gland. Strongest signals were detected in eye, kidneys and intestine.

Example 2: Generation and Analysis of Mice Comprising TDCDF1 Gene Disruptions

Targeting Construct. To investigate the role of Thymic Dendritic Cell-Derived Factor 1, disruptions in TDCDF1 genes were produced by homologous recombination. More particularly, as shown in Figure 4, a TDCDF1-specific targeting construct having the ability to disrupt or modify
5 TDCDF1 genes, specifically comprising SEQ ID NO:4 was created using as the targeting arms (homologous sequences) in the construct, the oligonucleotide sequences identified herein as SEQ ID NO:5 or SEQ ID NO:6.

Transgenic Mice. ES cells derived from the 129/SvJ x 129/Sv-CP mouse substrain were used to generate chimeric mice. F1 mice were generated by breeding with C57BL/6 females. F2
10 homozygous mutant mice were produced by intercrossing F1 heterozygous males and females.

Phenotypic Analysis. The transgenic mice were analyzed for phenotypic changes and expression patterns. The homozygous mice demonstrated at least one of the following phenotypes:

Expression: RT-PCR.

Total RNA was isolated from the organs or tissues from adult C57BL/6 wild type mice. RNA
15 was DNaseI treated, and reverse transcribed using random primers. The resulting cDNA was checked for the absence of genomic contamination using primers specific to non-transcribed genomic mouse DNA. cDNAs were balanced for concentration using HPRT primers. RNA transcripts were detected in all tissues analyzed: brain, cortex, subcortical region, cerebellum, brainstem, olfactory bulb, eye, heart, lung, liver, pancreas, kidneys, spleen, thymus, lymph nodes, bone marrow, skin, gall bladder,
20 urinary bladder, pituitary gland, adrenal gland, salivary gland, skeletal muscle, tongue, stomach, small intestine, large intestine, cecum, testis, epididymis, seminal vesicle, coagulating gland, prostate gland, ovary and uterus.

Example 3: Generation and Analysis of Mice Comprising TR2 Gene Disruptions

Targeting Construct. To investigate the role of Toll-Like Receptor 2, disruptions in TR2
25 genes were produced by homologous recombination. More particularly, as shown in Figure 6, a TR2-specific targeting construct having the ability to disrupt or modify TR2 genes, specifically comprising SEQ ID NO:7 was created using as the targeting arms (homologous sequences) in the construct, the oligonucleotide sequences identified herein as SEQ ID NO:8 or SEQ ID NO:9.

Transgenic Mice. ES cells derived from the 129/Sv-+P+Mgf-SLJ/J mouse substrain were
30 used to generate chimeric mice. F1 mice were generated by breeding with C57BL/6 females. F2 homozygous mutant mice were produced by intercrossing F1 heterozygous males and females.

Phenotypic Analysis. The transgenic mice were analyzed for phenotypic changes and expression patterns. The homozygous mice demonstrated at least one of the following phenotypes:

Expression: RT-PCR.

Total RNA was isolated from the organs or tissues from adult C57BL/6 wild type mice. RNA
35 was DNaseI treated, and reverse transcribed using random primers. The resulting cDNA was checked

for the absence of genomic contamination using primers specific to non-transcribed genomic mouse DNA. cDNAs were balanced for concentration using HPRT primers. RNA transcripts were detected in all tissues examined: brain, cortex, subcortical region, cerebellum, brainstem, olfactory bulb, eye, heart, lung, liver, pancreas, kidneys, spleen, thymus, lymph nodes, bone marrow, skin, gall bladder, urinary bladder, pituitary gland, adrenal gland, salivary gland, skeletal muscle, tongue, stomach, small intestine, large intestine, cecum, testis, epididymis, seminal vesicle, coagulating gland, prostate gland, ovary and uterus.

Example 4: Generation and Analysis of Mice Comprising nE3-like Gene Disruptions

Targeting Construct. To investigate the role of Ubiquitin Ligase E3s, disruptions in E3-like genes were produced by homologous recombination. More particularly, as shown in Figure 8, a E3-like-specific targeting construct having the ability to disrupt or modify E3-like genes, specifically comprising SEQ ID NO:10 was created using as the targeting arms (homologous sequences) in the construct, the oligonucleotide sequences identified herein as SEQ ID NO:11 or SEQ ID NO:12.

Transgenic Mice. ES cells derived from the 129/OlaHsd mouse substrain were used to generate chimeric mice. F1 mice were generated by breeding with C57BL/6 females. The resultant F1N0 heterozygotes were backcrossed to C57BL/6 mice to generate F1N1 heterozygotes. F2N1 homozygous mutant mice were produced by intercrossing F1N1 heterozygous males and females.

Phenotypic Analysis. The transgenic mice were analyzed for phenotypic changes and expression patterns. The homozygous mice demonstrated at least one of the following phenotypes:

20 Expression.

LacZ. Tissues of the transgenic animals were analyzed for expression of the target gene. Organs from one heterozygous male and one heterozygous female were frozen, sectioned (10 μ m), stained and analyzed for lacZ expression using X-Gal as a substrate for beta-galactosidase. Nuclear Fast Red was used for counterstaining.

25 Organs and tissues collected and frozen: brain, sciatic nerve, eye, Harderian glands, thymus, spleen, lymph nodes, bone marrow, aorta, heart, lung, liver, gallbladder, pancreas, kidney, urinary bladder, trachea, larynx, esophagus, thyroid gland, pituitary gland, adrenal glands, salivary glands, stomach, small and large intestines, tongue, skeletal muscle, skin and reproductive system.

In addition, the brain of the heterozygous female was analyzed for lacZ expression as wholemount. The dissected brain was cut longitudinally, fixed and stained using X-Gal as a substrate for beta-galactosidase. To stop the reaction the brain was washed in PBS and fixed in PBS-buffered formaldehyde.

30 Wild type control tissues were stained for X-gal to reveal background or signals due to endogenous beta-galactosidase activity. The following tissues show staining in the wild-type control sections and are therefore not suitable for X-gal staining: small and large intestines, stomach, vas
35

deferens and epididymis. It has been previously reported that these organs contain high levels of endogenous beta-galactosidase activity.

The results were as follows: LacZ (beta-galactosidase) expression was detectable in brain, lacrimal glands, Harderian glands, liver, kidney and seminal vesicles. Specific details of the staining pattern observed in each tissue are given below.

Brain. In brain wholemounts, faint signals were visible in ventricles. No X-Gal staining was detected on frozen sections.

Lacrimal Glands. Many cells in the lacrimal glands displayed strong X-Gal staining.

Harderian glands. X-Gal staining was detected in many cells.

Liver. Many hepatocytes displayed strong X-Gal staining.

Kidney. Many cells in the tubules of the cortex stained strongly. In the medulla, scattered lacZ expression was detected.

Male Reproductive System: Seminal Vesicles. Scattered X-Gal staining was detected in cells of the capsule.

RT-PCR. Total RNA was isolated from the organs or tissues from adult C57Bl/6 wild type mice. RNA was DNaseI treated, and reverse transcribed using random primers. The resulting cDNA was checked for the absence of genomic contamination using primers specific to non-transcribed genomic mouse DNA. cDNAs were balanced for concentration using HPRT primers. RNA transcripts were detected in all tissues analyzed: brain, cortex, subcortical region, cerebellum, brainstem, olfactory bulb, spinal cord, eye, Harderian glands, heart, lung, liver, pancreas, kidney, spleen, thymus, lymph nodes, bone marrow, skin, gallbladder, urinary bladder, pituitary gland, adrenal gland, salivary gland, skeletal muscle, tongue, stomach, small intestine, large intestine, cecum, testis, epididymis, seminal vesicle, coagulating gland, prostate gland, ovaries, uterus and white fat.

Example 5: Generation and Analysis of Mice Comprising CSPI Gene Disruptions

Targeting Construct. To investigate the role of Contrapsin-Like Serine Protease Inhibitors, disruptions in CSPI genes were produced by homologous recombination. More particularly, as shown in Figure 9, a CSPI-specific targeting construct having the ability to disrupt or modify CSPI genes, specifically comprising SEQ ID NO:13 was created using as the targeting arms (homologous sequences) in the construct, the oligonucleotide sequences identified herein as SEQ ID NO:14 or SEQ ID NO:15.

Transgenic Mice. ES cells derived from the 129/OlaHsd mouse substrain were used to generate chimeric mice. F1 mice were generated by breeding with C57BL/6 females. The resultant F1N0 heterozygotes were backcrossed to C57BL/6 mice to generate F1N1 heterozygotes. F2N1 homozygous mutant mice were produced by intercrossing F1N1 heterozygous males and females.

Phenotypic Analysis. The transgenic mice were analyzed for phenotypic changes and expression patterns. The homozygous mice demonstrated at least one of the following phenotypes:

Expression.

LacZ. Tissues of the transgenic animals were analyzed for expression of the target gene. Organs from one heterozygous male and one heterozygous female were frozen, sectioned (10 μ m), stained and analyzed for lacZ expression using X-Gal as a substrate for beta-galactosidase. Nuclear
5 Fast Red was used for counterstaining.

Organs and tissues collected and frozen: brain, sciatic nerve, eye, Harderian glands, thymus, spleen, lymph nodes, bone marrow, aorta, heart, lung, liver, gallbladder, pancreas, kidney, urinary bladder, trachea, larynx, esophagus, thyroid gland, pituitary gland, adrenal glands, salivary glands, stomach, small and large intestines, tongue, skeletal muscle, skin and reproductive system.

10 In addition, the brain of the heterozygous female was analyzed for lacZ expression as wholemount. The dissected brain was cut longitudinally, fixed and stained using X-Gal as a substrate for beta-galactosidase. To stop the reaction the brain was washed in PBS and fixed in PBS-buffered formaldehyde.

15 Wild type control tissues were stained for X-gal to reveal background or signals due to endogenous beta-galactosidase activity. The following tissues show staining in the wild-type control sections and are therefore not suitable for X-gal staining: small and large intestines, stomach, vas deferens and epididymis. It has been previously reported that these organs contain high levels of endogenous beta-galactosidase activity.

The results were as follows: LacZ (beta-galactosidase) expression was detected in skin, testis
20 and ovary. Specific details of the staining pattern observed in each tissue are given below.

Skin. A few epithelial cells of the epidermis showed X-Gal staining.

Skin of the Ear. Several epithelial cells of the epidermis showed strong X-Gal staining.

Male Reproductive System: Testis. Many spermatogenic cells in the seminiferous tubules expressed lacZ.

25 Female Reproductive System: Ovary. A few follicles showed faint X-Gal staining.

RT-PCR. Total RNA was isolated from the organs or tissues from adult C57Bl/6 wild type mice. RNA was DNaseI treated, and reverse transcribed using random primers. The resulting cDNA was checked for the absence of genomic contamination using primers specific to non-transcribed genomic mouse DNA. cDNAs were balanced for concentration using HPRT primers. RNA
30 transcripts were detected in olfactory bulb, liver, thymus, skin, urinary bladder, tongue, stomach and white fat. Highest levels of RNA transcripts were seen in skin and stomach.

Example 6: Generation and Analysis of Mice Comprising Abcd2 Gene Disruptions

Targeting Construct. To investigate the role of ABC transporter ATPase 2s, disruptions in Abcd2 genes were produced by homologous recombination. More particularly, as shown in Figure
35 11, a Abcd2-specific targeting construct having the ability to disrupt or modify Abcd2 genes, specifically comprising SEQ ID NO:16 was created using as the targeting arms (homologous

sequences) in the construct, the oligonucleotide sequences identified herein as SEQ ID NO:17 or SEQ ID NO:18.

Transgenic Mice. ES cells derived from the 129/OlaHsd mouse substrain were used to generate chimeric mice. F1 mice were generated by breeding with C57BL/6 females. The resultant
5 F1N0 heterozygotes were backcrossed to C57BL/6 mice to generate F1N1 heterozygotes. F2N1 homozygous mutant mice were produced by intercrossing F1N1 heterozygous males and females.

Phenotypic Analysis. The transgenic mice were analyzed for phenotypic changes and expression patterns. The homozygous mice demonstrated at least one of the following phenotypes:

Expression.

10 *RT-PCR.* Total RNA was isolated from the organs or tissues from adult C57Bl/6 wild type mice. RNA was DNaseI treated, and reverse transcribed using random primers. The resulting cDNA was checked for the absence of genomic contamination using primers specific to non-transcribed genomic mouse DNA. cDNAs were balanced for concentration using HPRT primers. RNA
15 transcripts were detected in brain, cortex, subcortical region, cerebellum, brainstem, olfactory bulb, eye, heart, lung, liver, pancreas, kidneys, spleen, thymus, bone marrow, skin, gall bladder, urinary bladder, pituitary gland, adrenal gland, salivary gland, skeletal muscle, tongue, stomach, small intestine, large intestine, cecum, testis, epididymis, seminal vesicle, coagulating gland, prostate gland, ovary and uterus.

20 As is apparent to one of skill in the art, various modifications of the above embodiments can be made without departing from the spirit and scope of this invention. These modifications and variations are within the scope of this invention.

46
CLAIMS

We claim:

1. A targeting construct comprising:
 - (a) a first polynucleotide sequence homologous to a target gene selected from the group consisting of: a CHD2-like gene, a TDCDF1 gene, a TR2 gene, an E3-like gene, a CSPI gene, and an Abcd2 gene;
 - (b) a second polynucleotide sequence homologous to the target gene; and
 - (c) a selectable marker.
2. The targeting construct of claim 1, wherein the targeting construct further comprises a screening marker.
3. A method of producing a targeting construct for a target gene selected from the group consisting of: a CHD2-like gene, a TDCDF1 gene, a TR2 gene, an E3-like gene, a CSPI gene, and an Abcd2 gene, the method comprising:
 - (a) obtaining a first polynucleotide sequence homologous to the target gene;
 - (b) obtaining a second polynucleotide sequence homologous to the target gene;
 - (c) providing a vector comprising a selectable marker; and
 - (d) inserting the first and second sequences into the vector, to produce the targeting construct.
4. A method of producing a targeting construct for a target gene selected from the group consisting of: a CHD2-like gene, a TDCDF1 gene, a TR2 gene, an E3-like gene, a CSPI gene, and an Abcd2 gene, the method comprising:
 - (a) providing a polynucleotide comprising a first sequence homologous to a first region of the target gene and a second sequence homologous to a second region of the target gene; and
 - (b) inserting a positive selection marker between the first and second sequences to form the targeting construct.
5. A cell comprising a disruption in a target gene selected from the group consisting of: a CHD2-like gene, a TDCDF1 gene, a TR2 gene, an E3-like gene, a CSPI gene, and an Abcd2 gene.
6. The cell of claim 5, wherein the cell is a murine cell.
7. The cell of claim 6, wherein the murine cell is an embryonic stem cell.
8. A non-human transgenic animal comprising a disruption in a target gene selected from the group consisting of: a CHD2-like gene, a TDCDF1 gene, a TR2 gene, an E3-like gene, a CSPI gene, and an Abcd2 gene.
9. A cell derived from the non-human transgenic animal of claim 8.
10. A method of producing a transgenic mouse, the method comprising:
 - (a) introducing the targeting construct of claim 1 into a cell;
 - (b) introducing the cell into a blastocyst;

- (c) implanting the resulting blastocyst into a pseudopregnant mouse, wherein said pseudopregnant mouse gives birth to a chimeric mouse; and
 - (d) breeding the chimeric mouse to produce the transgenic mouse.
11. A method of identifying an agent that modulates the expression of a target gene selected from the group consisting of: a CHD2-like gene, a TDCDF1 gene, a TR2 gene, an E3-like gene, a CSPI gene, and an Abcd2 gene, the method comprising:
- (a) providing a non-human transgenic animal comprising a disruption in the target gene;
 - (b) administering an agent to the non-human transgenic animal; and
 - (c) determining whether the expression of the disrupted target gene in the non-human transgenic animal is modulated.
12. A method of identifying an agent that modulates the function of a target gene selected from the group consisting of: a CHD2-like gene, a TDCDF1 gene, a TR2 gene, an E3-like gene, a CSPI gene, and an Abcd2 gene, the method comprising:
- (a) providing a non-human transgenic animal comprising a disruption in the target gene;
 - (b) administering an agent to the non-human transgenic animal; and
 - (c) determining whether the function of the disrupted target gene in the non-human transgenic animal is modulated.
13. A method of identifying an agent that modulates the expression of a target gene selected from the group consisting of: a CHD2-like gene, a TDCDF1 gene, a TR2 gene, an E3-like gene, a CSPI gene, and an Abcd2 gene, the method comprising:
- (a) providing a cell comprising a disruption in the target gene;
 - (b) contacting the cell with an agent; and
 - (c) determining whether expression of the target gene is modulated.
14. A method of identifying an agent that modulates the function of a target gene selected from the group consisting of: a CHD2-like gene, a TDCDF1 gene, a TR2 gene, an E3-like gene, a CSPI gene, and an Abcd2 gene, the method comprising:
- (a) providing a cell comprising a disruption in the target gene;
 - (b) contacting the cell with an agent; and
 - (c) determining whether the function of the target gene is modulated.
15. The method of claim 13 or claim 14, wherein the cell is derived from the non-human transgenic animal of claim 8.
16. An agent identified by the method of claim 11, claim 12, claim 13, or claim 14.
17. Phenotypic data associated with the transgenic mouse of claim 8, wherein the data is in a database.

TTGCCACATGGGAGAACAAGCTCTGTTTTAACTGTGAATGGTAGAACTGAGAACTATATTCTGGACACTCAACAT
 GGTGTCCAAGCATCTCTGGAATGTGCTGTTCAAAACCCACACCGAGGATGAAGATCTTCTCTGGTACAGAGAAGAT
 GGAATAGTAGATCTGAAAAATGGAAACAAAATCAACATCAGTTCTGTTTGTGTCTCTCCCATCAACGAAAGTGAC
 AATGGAGTCCGCTTTACCTGCAAGCTACAGCGGATCAGACGGTGTCCGTTACAGTAGTGCTGAACGTTACCTTT
 CCTCCTCTTCTAAGTGGCAATGGCTTCCAAACCGTGGAGGAAAACAGCGATGTAAGTTTGGTTTGCAACGTGAAA
 TCCAACCTCAAGCGCAAATGATGTGGTATAAAAACAACAGCGCCTTGGTTTGTAGAGAAAGGCCGTCACCACATC
 CACCAGACAAGGGAGTCTTTTCAGTTGTCCATCACCAAAGTCAAGAAATCTGACATTGGGACCTACAGCTGTATT
 GCCAGCTCATCTCTGAAG (SEQ ID NO:1)

FIGURE 1

Underlined = deleted in targeting construct

Bold = sequence flanking Neo insert in targeting construct

TTGCCACATGGGAGAACAAGCTCTGTTTTAACTGTGAATGGTAGAACTGAGAACTATATT
 CTGGACACTCAACATGGTGTCCAAGCATCTCTGGAATGTGCTGTTCAAAACCCACACCGAG
 GATGAAGATCTTCTCTGGTACAGAGAAGATGGAATAGTAGATCTGAAAAATGGAAACAAA
 ATCAACATCAGTTCTGTTTGTGTCTCTCCATCAACGAAAGTGACAATGGAGTCCGCTTT
 ACCTGCAAGCTACAGCGGATCAGACGGTGTCCGTTACAGTAGTGCTGAACGTTACCTTT
CCTCCTCTTCTAAGTGGCAATGGCTTCCAAACCGTGGAGGAAAACAGCGATGTAAGTTT
GTTTGC AACGTGAAATCCAACCTCAAGCGCAAATGATGTGGTATAAAAACAACAGCGCC
TTGGTTTGTAGAGAAAGGCCGTCACCACATCACCAGACAAGGGAGTCTTTTCAGTTGTCC
ATCACCAAAGTCAAGAAATCTGACATTGGGACCTACAGCTGTATTGCCAGCTCATCTCTG
AAG

FIGURE 2A

2/15

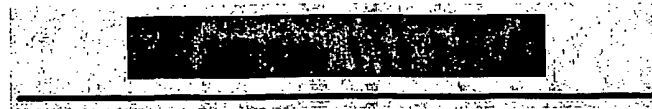
Gene Sequence Structure *

356 bp

Sequence Deleted

432 bp

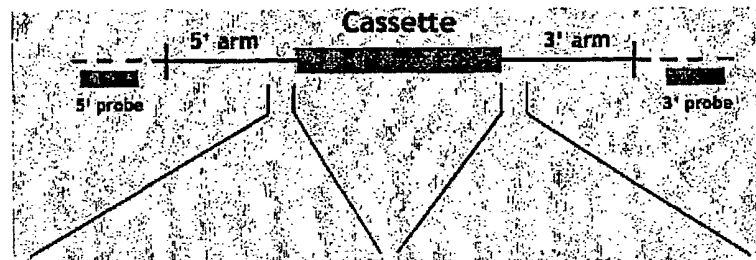
Size of EST: 543 bp



Targeting Vector* (genomic sequence)

Construct Number: 4033

LacZ-Neo



Arm Length:
5': 2.2 kb
3': 3.5 kb

Targeting Vector
Endogenous Locus

* Not drawn to scale

| | |
|--|---|
| <p>5' >TGTGTGTGTGTGTGTGTGTGT GTGTGTGTGTGTGTGTGTGTGTAAGGGT TTACGCCCTGAGAGAGATGGAAG GAAGAACTTGAATTGCTGACAATG GACTCACTACTGCATGCTTTTAAC GAATCTTGACTGCTTGAGTCTTAC AGTTCCTCCTCTTCTAAGTGGCAA TGGCTTCCAACCGTGGAGGAAAA CAGCGATGTAA<3' SEQ ID NO:2</p> | <p>5' >AAAGCCGTCACCACATCCAC CAGACAAGGGAGTCTTTTCAGTTG TCCATCACCAAGTCAAGAAATCT GACAATGGACCTACAGCTGTATT GCCAGCTCATCTCTGAAAATGGAG ACCATGGACTTCCACCTGCTTGT AAAGTAACTCTCTTCAAGCAATT GCTCAGTGTGGTGCATAAGTCAGG AAACAGGCTCT<3' SEQ ID NO:3</p> |
|--|---|

FIGURE 2B

3/15

CTGAAAAGAAAACAGCAACAATTCGAGCTGCTGTGACAGAGGGGAACAAGA
TGGCGGCGCCAAAGGGGAAGCTTTGGGTCCAGGCCCAACTGGGGCTCCCGC
CGCTGCTGCTGTTGACTATGGCGCTGGCCGGAGGCTCGGGGACTGCAGCGG
CCGAAGCCTTTGACTCGGTCCGGGAGACACAGCGTCCGTGCACCCGGCCCT
GTCAGCTGACCTACCCCTTGACACCTACCCGAAGGAAGAGGAGTTATACG
CATGCCAGAGAGGCTGCAGGCTGTTTTCAATTTGCCAGTTTGTGGATGATG
GGCTTGATTTAAATCGGACCAAGCTGGAATGTGAATCTGCGTGCACAGAAA
CATATTCCTAACCTGATGAGCAGTATGCTTGTGCATCTTGGCTGCCAGGATC
AGTGGCCATTTGCTGAACTGAGACAAGAACAAC'TCATGTCCCTGATGCCAA
GAATGCAGCTCCCTCTCCCTCTGACTCTGGTGAGGTCGTTCTGGAGTGACA
TGATGGACTCTGCACAGAGCTTCATAACCTCTTCATGGACTTTTTATCTTC
AAGCCGATGACGGAAAAATAGTTATATTCCAGTCTAAGCCAGAAAATTCAGT
ATGCACCGCAGTTGGAGCAGGAGCCTACAAACTTGAGAGAATCATCTTTAA
GCAAAATGTCCATCTGCAGATGAGAACTCACAAAGCACACAGGAACGACC
GTGAAGAGGAAGAAAGCGATGGCTTTTTAAGATGCTATCTCTTAACTCTG
GATGGATTTTAAACCACAACCCTTGTCTCTCGGTGATGGTGTGCTCTGGA
TCTGTTGTGCAGCTGTGCTACAGGTGTAGAACAGTATGTTCCCCCTGAGA
AGCTGAGTATCATTGGTCACTTCCAATTTATCAATCAACAAAACCTCACCA
CATACCCACCCTCCTCTCTTCTCATGTTAAGTCTCAGACTGAAGAACATG
AGGAGGCAGGGCCCTGCCACCAAGGTGAACCTTGCTCACTCAGAAATCT
AAGCTTTTTAAAAGAGTCGTGGACACATAAACTTCCATTCCTCATAACCT
TTTTAACATCCTTTCATTCGACATACCCCTTAACAAATCACTATAAAATCC
AAATAAAGTTACCAAACCTCTGTGAAGACTTTATTTGCTGTGACTTTACCTG
TATTTTCTAGTCAATTAAGATGGACATTTGGGTTGTATTTTTATTTTACTA
ATATCTGTAGCTACTTAGTTAGTTGCATTTGGTTTTGGTTTTTTCTCTCT
TCGCCAAATTCATGAGCTGATCATTGTGGCCCCGCCCTGCCATGCCCCC
CGTCAGTCATCTCACTTAATAACCGAAACCTTAGGGTGTGATGCTTGTGCC
CGGAAATGGCCTCCAAACCTGTCTGGGGATTATAGCACAAATGTTATTTAA
TGACACTACATTTTCAGTTGTATTGAATTGAAATCAITAAAATCTACTTGA
ATAATTATGTTCCAAAAAATAAAAAAAAAAAAAAAAAA (SEQ ID NO:4)

FIGURE 3

Underlined = deleted in targeting construct

Bold = sequence flanking Neo insert in targeting construct

CTGAAAAGAAAACAGCAACAATTCGAGCTGCTGTGACAGAGGGGAACAAGATGGCGGCGC
CAAAGGGGAAGCTTTGGGTCCAGGCCCAACTGGGGCTCCCGCCGCTGCTGCTGTTGACTA
TGGCGCTGGCCGGAGGCTCGGGGACTGCAGCGGCCGAAGCCTTTGACTCGGTCTGGGAG
ACACAGCGTCCGTGCACCGGGCCTGTGAGCTGACCTACCCCTTGCACACCTACCCGAAGG
AAGAGGAGTTATACGCATGCCAGAGAGGCTGCAGGCTGTTTTCAATTTGCCAGTTTGTGG
ATGATGGGCTTGATTTAAATCGGACCAAGCTGGAATGTGAATCTGCGTGCACAGAAACAT
ATTTCCCAACCTGATGAGCAGTATGCTTGTTCATCTTGGCTGCCAGGATCAGTGGCCATTTG
CTGAACTGAGACAAGAACAACCTCATGTCCCTGATGCCAAGAATGCAGCTCCTCTTCCCTC
TGACTCTGGTGAGGTCGTTCTGGAGTGACATGATGGACTCTGCACAGAGCTTCATAACCT
C'TTCATGGACTTTTTATCTTCAAGCCGATGACGGAAAAATAGTTATATTCCAGTCTAAGC
CAGAAATTCAGTATGCACCGCAGTTGGAGCAGGAGCCTACAAACTTGAGAGAATCATCTT
TAAGCAAATGTCCATCTGCAGATGAGAACTCACAAAGCACACAGGAACGACCGTGAAG
AGGAAGAAAGCGATGGCTTTTTAAGATGTCTATCTCTTAACTCTGGATGGATTTTAACCA
CAACCCTTGTCCCTCGGTGATGGTGTGCTCTGGATCTGTTGTGCAGCTGTTGCTACAG
GTGTAGAACAGTATGTTCCCCCTGAGAAGCTGAGTATCATTTGGTCACTTCCAATTTATCA
ATCAACAAAACCTCACACATACCCACCTCCTCCTCTTCTCATTTGTTAAGTCTCAGACTG
AAGAACATGAGGAGGCAGGGCCCCCTGCCACCAAGGTGAACCTTGCTCACTCAGAAATCT
AAGCTTTTTAAAAGAGTCGTGGACACATAAACTTCCATTCCTCATAACCTTTTTAACAT
CTTTTCATTCGACATACCCCTTAACAAATCACTATAAAATCCAAATAAAGTTACCAAACCT
CTGTGAAGACTTTATTTGCTGTGACTTTACCTGTATTTTTCTAGTCATTTAAGATGGACA
TTGGGTTGTATTTTTATTTACTAATATCTGTAGCTACTTAGTTAGTTGCATTTGGTTTTG
GTTTTTTTCCCTCTCTTCGCCAAATTCATGAGCTGATCATTTGGCCCCCGCCCCCTGCCAT
GCCCCCGTCCAGTCATCTCACTTAATAACCGAAACCTTAGGGTGTGATGCTTGTGCCCGG
AAATGGCCTCAAACGTCTCTGGGGATTATAGCACAAATGTTATTTAATGACACTACATT
TTCAGTTGTATTGAATTGAAATCATTTAAATCTACTTGAATAATTATGTTCCAAAAAAA
AAAAAAAAAAAAA

FIGURE 4A

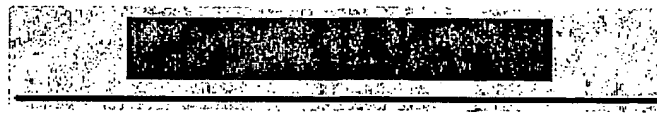
Gene Sequence Structure *

43 bp

Sequence Deleted

136 bp

Size of full-length
cDNA: 1513 bp



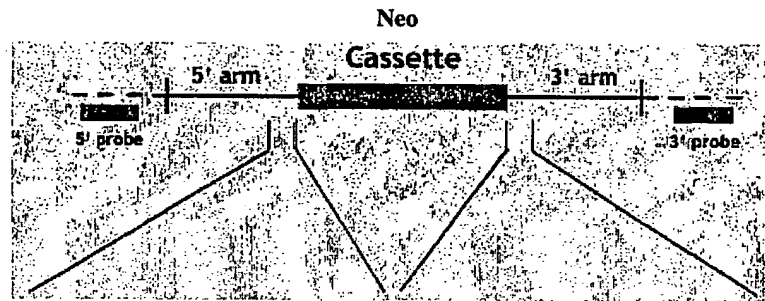
Targeting Vector* (genomic sequence)

Construct Number: 272

Arm Length:

5': 3.5 kb

3': 1.5 kb



* Not drawn to scale

| | |
|--|--|
| <p>5' >GCCGGGTCCCCAAGCCCGCCC ACGAGAGGTTTTCCGTCCCAGAGA CCGGCCAGGAGGCAGCCGATTGGG CGGACCGGGGCGGGCCGAGGAGC TC TCGGAGGCGGAAGCGCCGGCCG GAGGGCGGGGAAAAGGGGCTGGCC CGGAAGAAGGAGAAAGCCTGAAAA GAAAACAGCAACAATTCGAGCTGC TGTGACAGAGG<3' SEQ ID NO:5</p> | <p>5' >CTCGGGGACTGCAGCGGCCGA AGCCTTTGACTCGGTCTGGGAGA CACAGCGTCTGTACCAGGGCC TG TCAGCTGACCTACCCCTTGCACAC CTACCCGAAGGTAGGCAGGGCCCA TCCGTCCCGAGTCCCCTCATCCC TCGGTGCCCCAGCTCGCGTCGTTC CACCTCGGGGTCAGCGCCGGGGCC CTAATGGTGAA<3' SEQ ID NO:6</p> |
|--|--|

FIGURE 4B

GGCACGAGAGATGGCTAGTGGGCACGGGGAGCGGGCTGGAGGACTCCTAGGCTCCGGGCAGGCGGT
 CACTGGCAGGAGATGTGTCCGCAATCATAGTTTCTGATGGTGAAGGTTGGACGGCAGTCTCTGCGACC
 TAGAAGTGGAAAAGATGTCGTTCAAGGAGGTGCGGACTGTTTCCTTCTGACCAGGATCTTGTTTTCTGA
 GTGTAGGGGCTTCACTTCTCTGCTTTTCGTTTCATCTCTGGAGCATCCGAATTGCATCACCGGTCAGAA
 AACAACTTACCGAAACCTCAGACAAAGCGTCAAATCTCAGAGGATGCTACGAGCTCTTTGGCTCTTCT
 GGATCTTGGTGGCCATAACAGTCCCTCCTCAGCAAACGCTGTTCTGCTCAGGAGTCTCTGTCTATGTGAT
 GCTTCTGGGGTGTGTGATGGCCGCTCCAGGTCTTTCACCTCTATTCCTCCGGACTCACAGCAGCCAT
 GAAAAGCCTTGACCTGTCTTTCACCAAGATCACCTACATTTGGCCATGGTGACCTCCGAGCGTGTGCGA
 ACCTCCAGGTTCTGATTTTGAAGTCCAGCAGAATCAATACAATAGAGGGAGACGCCTTTTATTCTCTG
 GGCAGTCTTGAACATTTGGATTTGTCTGATAATCACCTATCTAGTTTATCTTCTCTCTGTTTCGGGCC
 CCTTTCCTCTTTGAAATACTTAAACTTAAATGGGAAATCCTTACCAGACACTGGGGGTAACATCGCTTT
 TTCCCAATCTCACAAATTTACAAACCTCAGGATAGGAAATGTAGAGACTTTCAGTGAGATAAGGAGA
 ATAGATTTTGTCTGGGCTGACTTCTCTCAATGAACTTGAAATTAAGGCATTAAGTCTCCGGAATTATCA
 GTCCCAAAGTCTAAAGTCGATCCGCGACATCCATCACCTGACTCTTCACTTAAAGCGAGTCTGCTTTCC
 TGCTGGAGATTTTTGTCAGATAATCTGAGTTCTGTGAGATAATTTAGAATAAGAGATACTAACTTGGCC
 AGGTTCCAGTTTTTACCCTGCCCCGTAGATGAAGTCAGCTCACCGATGAAGAAGCTGGCATTTCCGAGG
 CTCGGTTCTCACTGATGAAAGCTTTAACGAGCTCCTGAAGCTGTTGCGTTACATCTTGAAGTGTGCGG
 AGGTAGAGTTCGACGACTGTACCCCTCAATGGGCTCGGCGATTTCAACCCCTCGGAGTCAGACGTAGTG
 AGCGAGCTGGGTAAAGTAGAAACAGTCACTATCCGGAGGTTGCATATCCCCAGTCTATTTGTTTTA
 TGACCTGAGTACTGTCTATTCCTCCGGAGAAGGTGAAGCGAATCACAGTAGAGAACAGCAAGGCTCT
 TCCTGGTTCCCTGCTCGTTCCTCCAGCATTAAAATCATTAGAATCTTAGACCTCAGCGAAAATCTG
 ATGGTTGAAGAATATTTGAAGAACTCAGCCTGTAAGGGAGCCTGGCCTTCTCTACAAACCTTAGTTTT
 GAGCCAGAATCATTTGAGATCAATGCAAAAACAGGAGAGATTTTGCTGACTCTGAAAAACCTGACCT
 CCCTTGACATCAGCAGGAACACTTTTTATCCGATGCCCGACAGCTGTCACTGGCCGAAAAAGATGCGC
 TTCTTGAATTTGTCCAGTACAGGGATCCGGGTGGTAAAAACGTGCATTCCTCAGACGCTGGAGGTGTT
 GGATGTTAGTAACAACAATCTTGACTCATTTTCTTTGTTCTTGGCTCGGCTGCAAGAGCTCTATATTT
 CCAGAAATAAGCTGAAAACACTCCCAGATGCTTCGTTGTTCCCTGTGTTGCTGGTCATGAAAATCAGA
 GAGAATGCAGTAAGTACTTTCTCTAAAGACCAACTTGGTTCTTTTCCCAAACCTGGAGACTCTGGAAGC
 AGGCGACAACCACTTTGTTTGTCTCCTGCCAACTCCTATCCTTTACTATGGAGACGCCAGCTCTGGCTC
 AAATCCTGGTTGACTGGCCAGACAGCTACCTGTGTGACTCTCCGCTCGCTGCACGGCCACAGGCTT
 CAGGATGCCCCGCCCTCCGTCTTGGAAATGTACCAGGCTGCACGGTGTCTGGAGTCTGCTGTGCCCT
 TCTCTGTTGATCTTGTCTCGTAGGTGCCCTGTGCCACCATTTCACGGACTGTGGTACCTGAGAATGA
 TGTGGGCGTGGCTCCAGGCCAAGAGGAAGCCCAAGAAAGCTCCCTGCAGGGACGTTTGCTATGATGCC
 TTTGTTTCTACAGTGTGAGCAGGATTTCCCATTTGGGTGGAGAACCCTCATGGTCCAGCAGCTGGAGAATC
 TGACCCGCCCTTTAAGCTGTGTCTCCACAAGCGGGACTTCGTTCCGGGCAAATGGATCATTTGACAACA
 TCATCGATTCCATCGAAAAGAGCCACAAAACCTGTGTTCTGTTGCTTTCTGAGAATTCGTACGGAGCGAG
 TGGTGAAGTACGAACTGGACTTCTCCCACTTCCAGGCTCTTTGACGAGAACAACGACGGCCATCCT
 TGTTTTGCTGGAGCCATTGAGAGGAAAGCCATTCCCCAGCGCTTCTGCAAACGCGCAAGATAATGA
 ACACCAAGACCTACCTGGAGTGGCCCTTGGATGAAGGCCAGCAGGAAGTGTGTTGGGTAAATCTGAGA
 ACTGCAATAAAGTCTTAGGTTCTCCACCCAGTTCCTGACTTCCTTAAC TAAGGTC TTTGTGACACAAA
 CTGTAACAAAGTTTATAAGTAACATAGAATTGTATTATTGAGGATATTAAC TATGGGTTTTGTCTTGA
 ATACTGTTATATAAATATGTGACATCAGGAAAAAAAAAAAAAAAAAAAAA (SEQ ID NO:7)

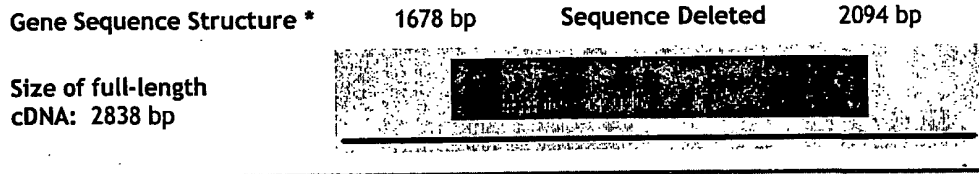
FIGURE 5

Underlined = deleted in targeting construct

Bold = sequence flanking Neo insert in targeting construct

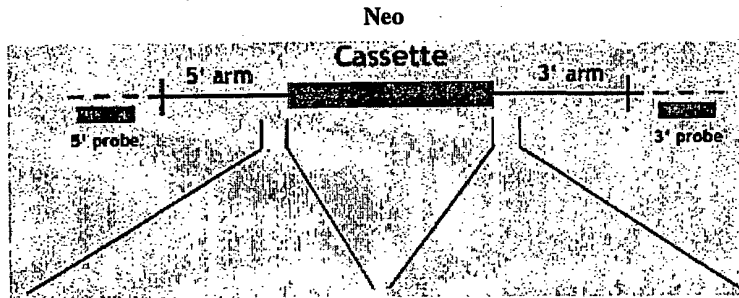
GGCACGAGAGATGGCTAGTGGGCACGGGGAGCGGGCTGGAGGACTCCTAGGCTCCGGG
CAGCGGTCAC TGGCAGGAGATGTGTCCGCAATCATAGTTTCTGATGGTGAAGGTGGAC
GGCAGTCTCTGCGACCTAGAAGTGAAAAGATGTCGTTCAAGGAGGTGCGGACTGTTTCC
TTCTGACCAGGATCTTGTTCGTAGTGTAGGGGCTTCACTTCTCTGCTTTTCGTTTCATCT
CTGGAGCATCCGAATTCATCACCAGGTCAGAAAACAACCTTACCGAAACCTCAGACAAAGC
GTCAAATCTCAGAGGATGCTACGAGCTCTTTGGCTCTTCTGGATCTTTGGTGGCCATAACA
GTCCCTTTCAGCAAACGCTGTTCTGCTCAGGAGTCTCTGTCATGTGATGCTTCTGGGGTG
TGTGATGGCCGCTCCAGGTCTTTCACTCTATTTCCCTCCGGACTCACAGCAGCCATGAAA
AGCCTTGACCTGTCCTTCAACAAGATCACCTACATTGGCCATGGTGACCTCCGAGCGTGT
GCGAACCTCCAGGTTCTGATTTTGAAGTCCAGCAGAATCAATACAATAGAGGGAGACGCC
TTTTATTCTCTGGGCAGTCTTGAACATTTGGATTTGTCTGATAATCACCTATCTAGTTTA
TCTTCTCTCTGGTTCGGGCCCCCTTTCCTCTTTGAAATACTTAAACTTAAATGGGAAATCCT
TACCAGACACTGGGGGTAACATCGCTTTTCCCAATCTCACAAATTTACAAACCTCAGG
ATAGGAAATGTAGAGACTTTCAGTGAGATAAGGAGAAATAGATTTTGTCTGGCTGACTTCT
CTCAATGAAC TGAATTAAGGCATTAAGTCTCCGGAATATCAGTCCCAAAGTCTAAAG
TCGATCCGCGACATCCATCACCTGACTCTTCACTTAAGCGAGTCTGCTTTCTGCTGGAG
ATTTTTGCAGATATCTGAGTCTGTGAGATATTTAGAACTAAGAGATACTAACTTGGCC
AGGTTCCAGTTTTCACCACTGCCGCTAGATGAAGTCAGCTCACCGATGAAGAAGCTGGCA
TTCCGAGGCTCGGTTCTCACTGATGAAAGCTTTAACGAGCTCCTGAAGCTGTTGCGTTAC
ATCTTGGAACTGTCGGAGGTAGAGTTCGACGACTGTACCTCAATGGGCTCGGCGATTTTC
AACCCCTCGGAGTCAGACGTAGTGAGCGAGCTGGGTAAAGTAGAAACAGTCACTATCCGG
AGGTTGCATATCCCCAGTCTATTTGTTTTATGACCTGAGTACTGTCTATTTCCCTCCTG
GAGAAGGTGAAGCGAATCACAGTAGAGAACAGCAAGGCTTTCTGCTGCTCCCTGCTCGTTC
TCCAGCATTTAAATCATTAGAATTTCTTAGACCTCAGCGAAAATCTGATGGTTGAAGAA
TATTTGAAGAACTCAGCCTGTAAGGGAGCCTGGCCTTCTCTACAAACCTTAGTTTTGAGC
CAGAATCATTGAGATCAATGCAAAAAACAGGAGAGATTTTGTGCTGACTTGAAAAACCTG
ACCTCCCTTGACATCAGCAGGAACACTTTTCATCCGATGCCCGACAGCTGTGAGTGGCCA
GAAAAGATGCGCTTCCTGAAATTTGTCCAGTACAGGGAATCCGGGTGGTAAAAACGTGCATT
CCTCAGACGCTGGAGGTGTTGGATGTTAGTAACAACAATCTTGACTCATTTTCTTTGTTT
TGCTTCCGCTGCAAGAGCTCTATATTTCCAGAAATAAGCTGAAAACACTCCCAGATGCT
TCGTTGTTCCCTGTGTTGCTGGTCAAGAAAATCAGAGAGAATGCAGTAAGTACTTTCTCT
AAAGACCAACTTGGTTCCTTTTCCCAAAC TGGAGACTCTGGAAGCAGGGCAGCAACCACTT
GTTGCTCCTGCGAACTCCTATCCTTTACTATGGAGACGCCAGCTCTGGCTCAAATCCTG
GTTGACTGGCCAGACAGCTACCTGTGTGACTCTCCGCTCGCCTGCACGGCCACAGGCTT
CAGGATGCCCCGGCCCTCCGCTCTTGGAAATGTCACCAGGCTGCCTGGTGTCTGGAGTCTGC
TGTGCCCTTCTCCTGTTGATCTTGTCTGCTAGGTGCCCTGTGCCACCAATTTCCACGGACTG
TGGTACCTGAGAAATGATGTGGGCTGGCTCCAGGCCAAGAGGAAGCCCAAGAAAAGCTCCC
TGCAGGGACGTTTGTATGATGCCCTTTGTTTCTACAGTGAGCAGGATTTCCATTTGGGTG
GAGAACCTCATGGTCCAGCAGCTGGAGAACTCTGACCCGCCCTTTAAGCTGTGTCTCCAC
AAGCGGACTTCGTTCCGGGCAATGGATCATTGACAACATCATCGATTCCATCGAAAAG
AGCCACAAAAC TGTGTTCTGCTGCTTTCTGAGAAC TTCGTACGGAGCGAGTGGTGCAGTAC
GAACTGGACTTCTCCCACTTCAGGCTCTTTGACGAGAAACAACGACGGGCCATCTTGT
TTGCTGGAGCCCATGAGAGGAAAGCCATTCCCCAGGCTTCTGCAACTGCGCAAGATA
ATGAACCAAGACCTACCTGGAGTGGCCCTTGGATGAAGGCCAGCAGGAAGTGTTTTGG
GTAATCTGAGAACTGCAATAAAGTCTTAGGTTCTCCACCCAGTTCCTGACTTCCTTAAC
TAAGGTCTTTGTGACACAAACTGTAACAAAGT TATAAGTAACATAGAAT TGTATTATTG
AGGATATTAAC TATGGGTTTTGTCTTGAATACGTTATATAAATATGTGACATCAGGAAA
AAAAAAAAAAAAAAAAAAAA

FIGURE 6A



Targeting Vector*
(genomic sequence)
Construct Number: 198

Arm Length:
5': 2.5 kb
3': 4 kb



— Targeting Vector
- - - Endogenous Locus
* Not drawn to scale

| | |
|--|---|
| <p>5' >TCTCTACAAACCTTAGTTTTG AGCCAGAATCATTTGAGATCAATG CAAAAAACAGGAGAGATTTTGCTG ACTCTGAAAAACCTGACCTCCCTT GACATCAGCAGGAACACTTCCAT CCGATGCCCGACAGCTGTCAGTGG CCAGAAAAGATGCGCTTCCTGAAT TTGTCCAGTACAGGGATCCGGGTG GTAAAAACGTG<3'</p> <p>SEQ ID NO:8</p> | <p>5' >AGTC TGCTGTGCCCTTCTCCT GTTGATCTTGCTCGTAGGTGCCCT GTGCCACCATTTCCACGGACTGTG GTACCTGAGAATGATGTGGGCGTG GCTCCAGGCCAAGAGGAAGCCAA GAAAGCTCCCTGCAGGGACGTTTG CTATGATGCCTTTGTTCCTACAG TGAGCAGGATCCCATTGGGTGGA GAACCTCATGG<3'</p> <p>SEQ ID NO:9</p> |
|--|---|

FIGURE 6B

9/15

CCTGTCC TGAGTGACCAACAGGTCGTGGAGCTGATTCCTGGGGGCACGGGCATCGTGGTGGA
 ATATGAGGATCGCTCCCGTTTCATCCAGCTGGTGC GGAAGGCCGGCTAGAGGAGAGCAAGGA
 GCAGGTGGCAGCCATGCAGGCAGGCTGTGCTGAAGGTGGTACCACAGGCTGTGCTGGACTTGC
 TGACCTGGCAAGAGTTGGAAAAGAAGGTGTGTGGGGACCAGAGGCACTGTGGATGCTCTGCC
 CAACGTCAC TCGGTTTGAGGACTTCGAACCATCTGACACACGCGTGCAGTACTTTTGGGAGG
 CTCTGAACAATTTACCAATGAGGACCGGAGCCGTTTCCTGCGCTTTGTACCCGGCCGCAGC
 GCCTCCCTGCTCGGATCTACATTTACCCAGACAAACTGGGCTATGAGACCACAGACGCCCTG
 CCTGAGTCTTCTACCTGCTCCAGCACGCTCTTCCTCCACACTATGCCAGTGCCAAGGTATG
 CGAGGAGAAGCTCCGCTACGCCGCTACAAC TGTGTGAGCTTTGTTCTTCCAGAGGCTTGGC
 ATGGACATCCGCTCAATACCCAGGCTTTACCAGCAGCCAAACTACTGTCTTCTGG (SEQ
 ID NO:10)

FIGURE 7

Underlined = deleted in targeting construct

Bold = sequence flanking Neo insert in targeting construct

CCTGTCC TGAGTGACCAACAGGTCGTGGAGCTGATTCCTGGGGGCACGGGCATCGTGGTG
 GAATATGAGGATCGCTCCCGTTTCATCCAGCTGGTGC GGAAGGCCGGCTAGAGGAGAGCA
 AGGAGCAGG**TGGCAGCCATGCAGGCAGGCTGTGCTGAAGGTGGTACCACAGGCTGTGCTGG**
ACTTGCTGACCTGGCAAGAGTTGGAAAAGAAGGTGTGTGGGGACCAGAGGCACTGTGGAT
GCTCTGCGCAACGTCAC TCGGTTTGAGGACTTCGAACCATCTGACACACGCGTGCAGTAC
TTTTGGGAGGCTCTGAACAATTTACCAATGAGGACCGGAGCCGTTTCCTGCGCTTTGTC
ACCGGCCGCAGCGCCTCCCTGCTCGGATCTACATTTACCCAGACAAACTGGGCTATGAGA
 CCACAGACGCCCTGCCTGAGTCTTCTACCTGCTCCAGCACGCTCTTCCTCCACACTATG
 CCAGTGCCAAGGTATGCGAGGAGAAGCTCCGCTACGCCGCTACAAC TGTGTGAGCTTTG
 TTCTTCCAGAGGCTTGGCATGGACATCCGCTCAATACCCAGGCTTTACCAGCAGCCAAAC
 TACTGTCTTCTGG

FIGURE 8A

10/15

Gene Sequence Structure *

200 bp

Sequence Deleted

321 bp

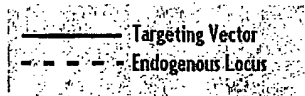
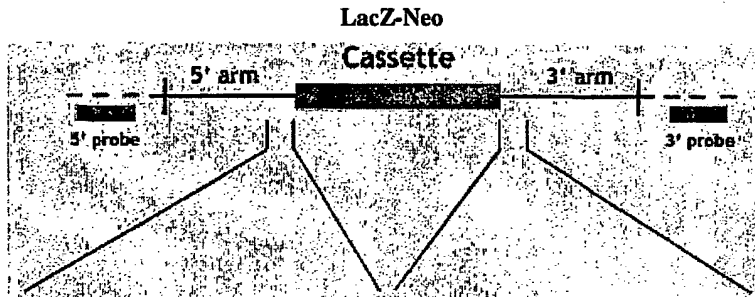
Size of EST: 613 bp



Targeting Vector*
(genomic sequence)

Construct Number: 3992

Arm Length:
5': 4 kb
3': 2.5 kb



* Not drawn to scale

| | |
|--|---|
| <p>5' >CAATATCCACCCTGATGCTGG CTTCCATGACTGCACCCATTCTT GCTGGTCAGTGAGTACAGAGATGC CCCAGTGTACGCTGAAAGTAGTGT CCAACTTTCCCAGGCAGGCTTCTT CCTCCTCCCTAGGTGGCAGCCATG CAGGCAGGTCTGCTGAAGGTGGTA CCACAGGCTGTGCTGGACTTGCTG ACCTGGCAAGA<3' SEQ ID NO:11</p> | <p>5' >TTCACCAATGAGGACCGGAGC CGTTTCCTGCGCTTGTACCCGGC CGCAGCCGCCCTCCCTGCTCGGATC TACATTTACCCAGACAAACTGGGG TGAGTATGAGTACAGAGAGGGGCA GAAGGTGCCCCACACACACATG CACGCATACCGCTCACTCCACAAC TTATAATCCTTTGCCCTTCCCACC TTGCCGCTATG<3' SEQ ID NO:12</p> |
|--|---|

FIGURE 8B

11/15

AGCAGCTTCAATTGGCCACACTCTGCCTCTTCAGGAGGCCTGGAGGTATAAATAATAAAACAAACCAGGCAGGCC
 CGTTTGGCCCTGTGGATGGCGCCTGGACCCACTGATAATCCCCAAGGACAATGACTCGCATGCTGGACCTGGGCC
 TGTTTCTGGCAGGCCCTCCTCACTGTGAAAGGTCCTTGCAGGACAGAGATGCTCCAGATATGTATGATTCTCCAG
 TCAGGGTCCAAGAGTGGAGGGGCAAGAAGGATGCCCGGCAGCTTGCTCGACACAACATGGAATTTGGCTTCAAGC
 TACTGCAGAGGCTGGCCAGCAATAGCCCTCAGGGAAACATCTTCCATCCCCACTGAGCATCTCCACTGCCTTCT
 CCATGCTGTCTCTCGGGGCCAGAACTCCACACTGGAGGAAATCCGGGAAGGCTTCAACTTCAAAGAGATGTCAA
 ACTGGGACGTCCATGCGGCTTTCCACTACCTTCTCCACAAGCTCAACCAGGAGACGGAGGACACAAAGATGAACT
 TAGGCAATGCTTTATTTATGGATCAGAACTGAGGCCCAACAGAGGTTCTTGAACCTGGCTAAAATGTGTATGATGC
 GACATGGTGCCATTACTTCCAGACTT (SEQ ID NO:13)

FIGURE 9

Underlined = deleted in targeting construct

Bold = sequence flanking Neo insert in targeting construct

AGCAGCTTCAATTGGCCACACTCTGCCTCTTCAGGAGGCCTGGAGGTATAAATAATAAAA
 CAAACCAGGCAGGCCCGTTTGGCCCTGTGGAT**GGCGCCTGGACCCACTGATAATCCCCAA**
GGACAATGACTCGCATGCTGGACCTGGGCCTGTTTCTGGCAGGCCCTCCTCACTGTGAAAG
GTCTCCTGCAGGACAGAGATGCTCCAGATATGTATGATTCTCCAGTCAGGGTCCAAGAGT
GGAGGGGCAAGAAGGATGCCCGGCAGCTTGCTCGACACAACATGGAATTTGGCTTCAAGC
TACTGCAGAGGCTGGCCAGCAATAGCCCTCAGGGAAACATCTTCCATCCCCACTGAGCA
TCTCCACTGCCTTCTCCATGCTGTCTCTCGGGGCCAGAACTCCACACTGGAGGAAATCC
GGGAAGGCTTCAACTTCAAAGAGATGTCAAAGTGGGACGTCCATGCGGCTTCCACTACC
TTCTCCACAAGCTCAACCAGGAGACGGAGGACACAAAGATGAACTTAGGCAATGCTTTAT
TTATGGATCAGAACTGAGGCCCAACAGAGGTTCTTGAACCTGGCTAAAATGTGTATGATGC
GACATGGTGCCATTACTTCCAGACTT

FIGURE 10A

12/15

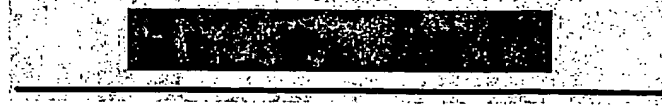
Gene Sequence Structure *

142 bp

Sequence Deleted

366 bp

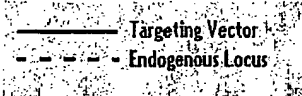
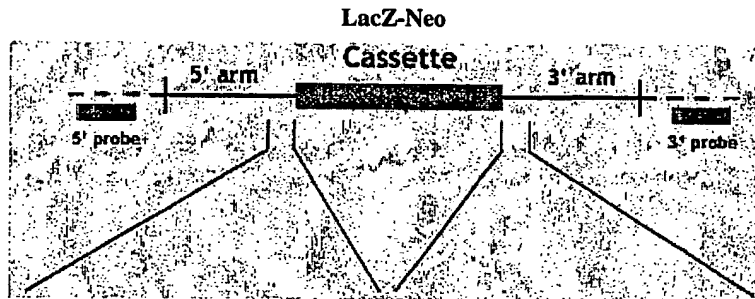
Size of EST: 626 bp



Targeting Vector*
(genomic sequence)

Construct Number: 2995

Arm Length:
5': 2.7 kb
3': 2 kb



* Not drawn to scale

| | |
|---|--|
| <p>5' >GGAGATGACAAAGATTCAATG CCATCAAGACTCTCAAAGCACTCG GCTGTGGTGTTCAGCACCTACAG CCGCTGTCTTCTCCCATGGGTC'TT CCTGGTGGCTAAGATGGGAGGGAG CCATGGGGTCC'TCCCTTCTTGCT TTAACCTCAGATCTGCTGCTCAAC AGGATAATCCCAAGGACAATGAC TCGCATGCTGG<3'</p> <p>SEQ ID NO:14</p> | <p>5' >CTGCCCTTCCATGCTGTCTC TCGGGGCCCAGAACTCCACACTGG AGGAAATCCGGGAGGGCTTCAACT TCAAAGAGATGTCAAACAGGGACG TCCATGGGGCTTTCCTACTACCTTC TCCACAAGCTCAACCAGGAGACGG AGGACACAAAGATGAACTTAGGCA ATGCTTTATTTATGGATCAGAAGC TGAGGCCCCAA<3'</p> <p>SEQ ID NO:15</p> |
|---|--|

FIGURE 10B

CGACAGCAACGGCGCCCCGGAGGATGCGGTGGAGTTTGTGCTTTGCTGCATCCGTCACTGAAGAACAAAATAAG
AGTAAAGGAGACAAGCTGCAGAGCATGGGAGGCTGTGGGGTCTTCTGAAACCTTTGCTGGGCCTTCCGCGGAGCA
TGAGCTTTTAAAACGAATTCTTTTCAAAGAAACCCATTGTGTAGCTGGAAAAATGATACACATGCTAAATGCAG
CAGCCTATCGGGTAAAATGGACCAGATCCGGTGTCTGCTAAAAGGGCTGCCCTGCCCTGGTGGCTCGGGCATATGCTC
TGAAAACCCCTATCCCATCATTTGGCAAGCGTTTAAAGCAGCCTGGCCACAGGAAGGCAAAGCAGAAGCTTACT
CGCTGCAGAGAACAGAGAAATACCTGCATTGCACGGAGATCATCTGTAAAAACCTGCGCCGGGACTAAATGCAG
CTTTTTTCAAACAGCTACTAGAACCTCGGAAAATCCTCTTTCCAAAACCTTGACCCTGAAACGGGGTGGCTCT
GCCTCCACTCGGTGGCTCTAATCTCAAGAACATTTCTCTCTATTTATGTGGCTGGTCTGGATGGGAAAATCGTGA
AAAGCATCGTGGAAAAGAAGCCTCGGACTTTCATCATCAAATTAATCAAGTGGCTTATGATGCTATCCCTGCTA
CCTTTGTCAACAGTGTATCAGGTACCTGGAATGCAACTGGCATTGGCCTTTAGAAGCTGCTTAGTAGACCATG
CCTATGAGACCTATTTTCGCAAATCAGACTTATTATAAGGTGATAAATATGGATGGGAGGCTGGCAAACCTGACC
AGTCTCTTACCGAAGACATTATGATGTTCTCGCAATCTGTGGCTCACCTGTATTCCAACCTTACCAAACCTATTT
TAGATGTCATTTAACCCTCTATCTCATCCGGACAGCTACATCCAGAGGAGCAAGCCCTATAGGGCCCACCC
TGTTAGCAGGACTTGTGCTGATGCCACTGCTAAAGTACTGAAAGCTTGCTCGCCCAAATTTGGTTCGCTGGTGG
CTGAAGAAGCCCACAGGAAAGGCTACCTGCGGTATGTCCACTCCCGAATCATAGCCAATGTAGAAGAAATTCGCT
TCTACAGAGGACATAAGGTAGAAATGAAGCAGCTGCAGAAATGTTACAAGGCTTTAGCTTACCAGATGAACCTGA
TTTTATCCAAACGTTTATGGTACATCATGATAGAACAATTTCTTGATGAAGTATGTGTGGAGCAGCTGTGGACTAA
TTATGGTGGCTATAACCCATTATCACTGCAACGGGCCTTTCGAGATGGTGTCTGGAGGATGTTCCAAAGCAGGCTA
TGGTTAGCGATCGGACAGAGGCTTACCCTGCCCCGAACTTACTGGCTCTGGAGCTGATGCAATTGAAAGGA
TTATGCTTTCATACAAAGAGATCACTGAACTAGCAGGTTATACTGCTAGAGTATAACAATATGTTCTGGGTCTTCG
ATGAAGTGAAGAGAGGCATTTATAAGAGAACTGTCACTCAGGAACCTGAAAACCATAGCAAGCGTGGAGGTAACC
TGAAACTACCCCTCAGCGACACCCGTCATCAAAGGAACAGTTATTGATGTGGATCATGGAATCATTGTGAAA
ATGTTCCCATAAATACACCAGCGGGCAAGTGGTGGCTTCCAGGCTAAACTTCAAAGTGGAGAAGGGATGCATC
TCTTGATAACTGGTCCCAACGGTGTGGGAAAAGCTCTCTCTTCAGAATCTTAAAGCGGCTGTGGCTGTGTATG
AAGGAGTCCCTTATAAAACCGCCTCCCAACATATGTTCTATATTCCACAGAGGCCATACATGCTCTTGGAAAGTC
TCCGGGATCAAGTCATTTACCTGACTCAGCGGATGCATGCGTGAGAAAAGGTTACACTGACCAAGACCTAGAAC
GCATCCTGCACAGCGTGCACCTCTACCACATAGTTCAAAGAGAAGGAGGATGGGATGCAGTCATGGACTGGAAG
ATGTCCCTTCCGGAGGGGAGAAGCAGAGAATGGGCATGGCGCGGATGTTTTACCATAAACCAGATATGCATTGC
TGGATGAATGTACCAGTGCCGTGAGCATCGACGTTGAAGGAAAGATATTTAGGCTGCTATTGGGGCTGGGATTT
CCCTACTCTCCATAACACACAGGCTTCTCTGTGGAAATACCACACTCATCTATTACAATTCGATGGCGAAGGAG
GCTGGCGCTTTGAACAGTTGGACTGCTATCCGTTTAAAGTGTAGTGAAGAAAAGCAAAGTTGGAGTCCGAGC
TCGCTGGAATTTCCAAAATGCAACAGAGACTCAACGAACTATGCAAAATTTGGGGGAAGACTCGGTGCTGAAA
CAATCCAACTCCAGAAAAGACATCTAATTTATCTTGACATGTTTTCAGTTACCTTCTAGGTGAAGCCTCAGAG
ACTCTCTCTTTACTGCATGCAGTATGTTAAGCTAAGTGCAGAGAAAGCAAGCCGC SEQ ID NO:16

FIGURE 11

Underlined = deleted in targeting construct

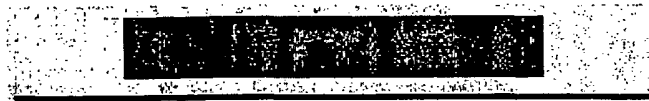
Bold = sequence flanking Neo insert in targeting construct

CGACAGCAACGGCGCCCCCGGAGGATGCGGTGGAGTTTGTGCTTTGCTGCATCCGTC
 ACTGAAGAACAAAATAAGAGTAAAGGAGACAAGCTGCAGAGCATGGGAGGCTGTGGGGTCTTC
 TGAAACCTTTGCTGGGCTTTCCGCGGAGCATGAGCTTTTAAAACGAATCTTTTCAAAGA
AACCCATTTGTGTAGCTGGAAAAATGATACACATGCTAAATGCAGCAGCCTATCGGGTGA
AATGGACCAGATCCGGTGCTGCTAAAAGGGCTGCCTGCCTGGTGGCTGCGGCATATGCTC
TGAAAACCTCTATCCCATCATTTGGCAAGCGTTTAAAGCAGCCTGGCCACAGGAAGGCAA
AAGCAGAAGCTTACTCGCCTGCAGAGAACAGAGAAAATACTGCATTGCACGGAGATCATCT
GTAAGAAAACCTGCGCCGGGACTAAAATGCAGCTTTTTTCAAACAGCTACTAGAACTTCGGA
AAATCCTCTTTCCAAAACTTGTGACCACCTGAAAACGGGGTGGCTCTGCCTCCACTCGGTGG
CTCTAATCTCAAGAACATTTCTCTCTATTTATGTGGCTGGTCTGGATGGGAAAATCGTGA
AAAGCATCGTGGAAAAGAAGCCTCGGACTTTTCATCATCAAATTAATCAAGTGGCTTATGA
TTGCTATCCCTGCTACCTTTGTCAACAGTGCATCAGGTACCTGGAATGCAAACCTGGCAT
TGGCCTTTAGAACTCGCTTAGTAGACCATGCCATGAGACCATTTTCGCAAATCAGACTT
ATTATAAGGTGATAAATATGGATGGGAGGCTGGCAAACCTGACCAGTCTCTTACCGAAG
ACATTATGATGTTCTCGCAATCTGTGGCTCACCTGTATTTCAAACCTTACCAAACCTATTT
TAGATGTCATTTCTAACCTCCTATACTCTCATCCGGACAGCTACATCCAGAGGAGCAAGCC
CTATAGGGCCCCACCTGTTAGCAGGACTTGTCTGTATGCCACTGCTAAAGTACTGAAAG
 CTTGCTCGCCCCAAATTTGGTTCGCTGGTGGCTGAAGAAGCCCACAGGAAAGGCTACCTGC
 GGTATGTCCACTCCCGAATCATAGCCAATGTAGAAGAAATGGCTTCTACAGAGGACATA
 AGGTAGAAAATGAAGCAGCTGCAGAAAATGTTACAAGGCTTTAGCTTACCAGATGAACCTGA
 TTTTATCCAAAACGTTTATGGTACATCATGATAGAACAATTTCTTGATGAAGTATGTGTGGA
 GCAGCTGTGGACTAATTATGGTGGCTATACCCATTATCACTGCAACGGGCTTTGCAGATG
 GTGATCTGGAGGATGGTCCAAAGCAGGCTATGGTTAGCGATCGGACAGAGGCCCTTACCA
 CTGCCCCGGAACCTFACTGGCCTCTGGAGCTGATGCAATTTGAAAGGATTATGTCTTCATACA
 AAGAGATCACTGAACTAGCAGGTTATACTGCTAGAGTATAACAATATGTTCTGGGTCTTCG
 ATGAAGTGAAGAGAGGCATTTATAAGAGAAGTGTCACTCAGGAACCTGAAAACCATAGCA
 AGCGTGGAGGTAACCTGGAACCTACCCCTCAGCGACACCCTGGCCATCAAAGGAACAGTTA
 TTGATGTGGATCATGGAATCATTTGTGAAAATGTTCCATAATTACACCAGCGGGCGAAG
 TGGTGGCTTCCAGGCTAACTTCAAAGTGAAGAAGGGATGCATCTCTTGATAACTGGTC
 CCAACGTTGTGGGAAAAGCTCTCTCTTCCAGAACTTAAAGCGGGCTGTGGCCTGTGTATG
 AAGGAGTCCTTTATAAACCGCCTCCCCAACATATGTTCTATATTCCACAGAGGCCATACA
 TGTCTCTTGGAAAGTCTCCGGGATCAAGTCATTTACCCTGACTCAGCGGATGACATGCGTG
 AGAAAGGTTACACTGACCAAGACCTAGAACGCATCCTGCACAGCGTGCACCTTACCACA
 TAGTTCAAAGAGAAGGAGGATGGGATGCAGTCACTGGACTGGAAAGATGTCTTTCCGGAG
 GGGAGAAGCAGAGAATGGGCATGGCGGGATGTTTACCATAAAACCGAAGTATGCATTGC
 TGGATGAATGTACCAGTCCCGTGAGCATCGACGTTGAAGGAAAAGATATTTTCCAGGCTGCTA
 TTGGGGCTGGGATTTCCCTACTCTCCATAACACACAGGCCCTTCTGTGGAAATACCACA
 CTCATCTATTACAATTCGATGGCGAAGGAGGCTGGCGCTTTGAACAGTTGGACACTGCTA
 TCCGTTTAAACGTTGAGTGAGGAAAAGCAAAAGTTGGAGTCCGACGCTCGCTGGAATTTCCA
 AAATGCAACAGAGACTCAACGAACATGCAAAAATCTGGGGGAAGACTCGGTGCTGAAAA
 CAATCCAAACTCCAGAAAAGACATCCTAATTTATCTTGACATGTTTTCAGTTACCTTCTA
 GGTGAAGCCTCAGAGACTCTCTCTTTACTGCATGCAGTATGTTAAGCTAAGTGCAGAGAA
 AGCAAGCCGGC

FIGURE 12A

Gene Sequence Structure * 489 bp Sequence Deleted 528 bp

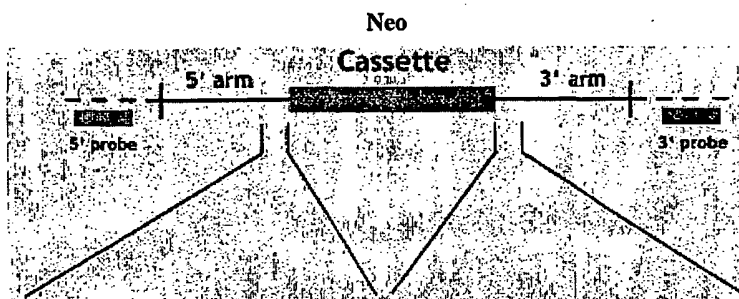
Size of full-length
cDNA: 2531 bp



Targeting Vector*
(genomic sequence)

Construct Number: 4773

Arm Length:
5': 1.3 kb
3': 3 kb



— Targeting Vector
- - - - Endogenous Locus
* Not drawn to scale

| | |
|--|--|
| <p>5' >CGGCATATGCTCTGAAAACCC TCTATCCCATCATTGGCAAGCGTT TAAAGCAGCCTGGCCACAGGAAGG CAAAAGCAGAAGCTTACTCGCCTG CAGAGAACAGAGAAATACTGCATT GCACGGAGATCATCTGTA AAAAAC CTGCGCCGGGACTAAATGCAGCTT TTTTCAAACAGCTACTAGAACTTC GGAAAATCCTC<3' SEQ ID NO:17</p> | <p>5' >TCCACTCGGTGGCTCTAATCT CAAGAACATTTCTCTATTTATG TGGCTGGTCTGGATGGGAAAATCG TGAAAAGCATCGTGGAAAAGAAGC CTCGGACTTTCATCATCAAATTA TCAAGTGGCTTATGATGCTATCC CTGCTACCTTTGTCAACAGTGCTA TCAGGTACCTGGAATGCAAACCTGG CATTGGCCTTT<3' SEQ ID NO:18</p> |
|--|--|

FIGURE 12B

SEQUENCE LISTING

<110> Deltagen, Inc.

<120> TRANSGENIC MICE CONTAINING TARGETED GENE
DISRUPTIONS

<130> 136 PCT

<150> US 60/255,983

<151> 2000-12-13

<150> US 60/255,972

<151> 2000-12-13

<150> US 60/255,968

<151> 2000-12-13

<150> US 60/256,232

<151> 2000-12-13

<150> US 60/254,314

<151> 2000-12-08

<150> US 60/251,819

<151> 2000-12-06

<160> 18

<170> FastSEQ for Windows Version 4.0

<210> 1

<211> 543

<212> DNA

<213> Mus musculus

<400> 1

```

ttgccacatg ggagaacaag ctctgtttta actgtgaatg gtagaactga gaactatatt 60
ctggacactc aacatgggtg ccaagcatct ctggaatgtg ctgttcaaaa ccacaccgag 120
gatgaagatc ttctctggta cagagaagat ggaatagtag atctgaaaaa tggaaacaaa 180
atcaacatca gttctgtttg tgtctctccc atcaacgaaa gtgacaatgg agtccgcttt 240
acctgcaagc tacagcgcga tcagacggtg tccggtacag tagtgctgaa cgttaccctt 300
cctcctcttc taagtggcaa tggcttccaa accgtggagg aaaacagcga tgtaagtttg 360
gtttgcaacg tgaatccaa ccctcaagcg caaatgatgt ggtataaaaa caacagcgcc 420
ttggttttag agaaaggccg tcaccacatc caccagacaa gggagtcttt tcagttgtcc 480
atcaccaaag tcaagaaatc tgacattggg acctacagct gtattgccag ctcctctctg 540
aag                                                                                   543

```

<210> 2

<211> 200

<212> DNA

<213> Artificial Sequence

<220>

<223> Targeting vector

<400> 2

```

tgtgtgtgtg tgtgtgtgtg tgtgtgtgtg tgtgtaaggg gtgaattacg ccctgagag 60
agatggaagg aagaacttga attgctgaca atggactcac tactgcatgc ttttaacgaa 120
tcttgactgc ttgagtctta cagttcctcc tcttctaagt ggcaatggct tccaaaccgt 180

```


ggaggaaaac agcgatgtaa

200

<210> 3
<211> 200
<212> DNA
<213> Artificial Sequence

<220>
<223> Targeting vector

<400> 3
aaaggcgc accacatcca ccagacaagg gagtcttttc agttgtccat caccaaagtc 60
aagaaatctg acaatgggac ctacagctgt attgccagct catctctgaa aatggagacc 120
atggacttcc acctgcttgt taaaggtaac tctcttcaag caattgctca gtgtgggtgca 180
taagtcagga aacaggctct 200

<210> 4
<211> 1513
<212> DNA
<213> Mus musculus

<400> 4
ctgaaaagaa aacagcaaca attcgagctg ctgtgacaga ggggaacaag atggcggcgc 60
caaaggggaa gctttgggct caggcccaac tggggctccc gccgctgctg ctggttgacta 120
tggcgctggc cggaggctcg gggactgcag cggccgaagc ctttgactcg gtcctgggag 180
acacagcgtc cgtgcaccgg gcctgtcagc tgacctacc cttgcacacc taccgaagg 240
aagaggagtt atacgcatgc cagagaggct gcaggctgtt ttcaatttgc cagtgtgtgg 300
atgatgggct tgatttaa atcgaccaagc tggaaatgga atctgcgtgc acagaaacat 360
attccaacc tgatgagcag tatgcttgtc atcttggctg ccaggatcag tggccattg 420
ctgaactgag acaagaaca ctcatgtccc tgatgccaag aatgcagctc ctcttcctc 480
tgactctggt gaggctgctc tggagtgaca tgatggactc tgcacagagc ttcataacct 540
cttcatggac tttttatctt caagccgatg acggaaaaat agttatattc cagtctaagc 600
cagaaattca gtatgcaccg cagttggagc aggagcctac aaacttgaga gaatcatctt 660
taagcaaaat gtcctatctg cagatgagaa actcacaagc acacaggaac gaccgtgaag 720
aggaagaaag cgatggcctt ttaagatgtc tatctcttaa ctctggatgg attttaacca 780
caacccttgt cctctcgggt atgggtgtgc tctggatctg ttgtgcagct gttgctacag 840
gtgtagaaca gtatgttccc cctgagaagc tgagtatcat tggctacttc caatttatca 900
atcaacaaaa cctcaccaca taccacctc ctctcttct cattgttaag tctcagactg 960
aagaacatga ggaggcagg cccctgcccc ccaaggtgaa ccttgctcac tcagaaatct 1020
aagcttttta aaagatcgt ggacacataa acttccattc ttcatacacc tttttaacat 1080
cctttcattc tacatacccc ttaacaatc actataaaat ccaataaag ttaccaaact 1140
ctgtgaagac tttatttgct gtgactttac ctgtattttt ctagtcatth aagatggaca 1200
ttgggttcta tttttatttt actaatatct gtactactt agttagtgtc attgggtttg 1260
gtttttttcc tctcttggcc aaattctatg agctgatcat tgtggccccg cccctgcat 1320
gcccccgctc agtcatctca cttaataacc gaaaccttag ggtgtgatgc ttgtgcccgg 1380
aatggcctc caaactgtcc tggggattat agcacaatg ttatttaatg aactacatt 1440
ttcagttgta ttgaattgaa atcattaaaa tctacttgaa taattatggt ccaaaaaaaa 1500
aaaaaaaaaaa aaa 1513

<210> 5
<211> 200
<212> DNA
<213> Artificial Sequence

<220>
<223> Targeting vector

<400> 5
gccgggtccc caagcccgc cagcagaggt tttccgtccc agagaccggc caggaggcag 60
ccgattgggc ggaccggggc ggggccgagg agctctcgga ggcggaagcg ccggccggag 120
gggcggggaa aggggctggc ccggaagaag gagaaagcct gaaaagaaaa cagcaacaat 180

tcgagctgct gtgacagagg

200

- <210> 6
- <211> 200
- <212> DNA
- <213> Artificial Sequence

- <220>
- <223> Targeting vector

<400> 6
 ctcggggact gcagcggccg aagcctttga ctcggctctg ggagacacag cgtcctgtca 60
 ccgggcctgt cagctgacct accccttgca cacctaccg aaggtaggca gggcccatcc 120
 gtcccagatt cccctcatcc ctcgggtgcc cagctcgcgt cgttccacct cggggtcagc 180
 gccggggccc taatgggtgaa 200

- <210> 7
- <211> 2838
- <212> DNA
- <213> Mus musculus

<400> 7
 ggcacgagag atggctagtg ggcacgggga ggggggctg gaggactcct aggctccggg 60
 caggcgggca ctggcaggag atgtgtccgc aatcatagtt tctgatgggtg aaggttggac 120
 ggcagtctct gcgacctaga agtggaaaag atgtcgttca aggaggtgcg gactgtttcc 180
 ttctgaccag gatcctgttt ctgagtgtag gggcttcaact tctctgcttt tcgttcatct 240
 ctggagcatc cgaattgcat caccggctcag aaaacaactt accgaaacct cagacaaagc 300
 gtcaaatctc agaggatgct acgagctctt tggctcttct ggatccttggg gcccataaca 360
 gtctctctca gcaaacgctg ttctgctcag aggtctctgt catgtgatgc ttctgggggtg 420
 tgtgatggcc gctccaggtc ttccacctct attcctctcg gactcacagc agccatgaaa 480
 agccttgacc tgtctttcaa caagatcacc tacattggcc atgggtgacct ccgagcgtgt 540
 gcgaacctcc aggttctgat tttgaagtcc agcagaatca atacaataga gggagacgcc 600
 ttttattctc tgggcagtct tgaacatttg gatttgtctg ataatcacct atctagttta 660
 tcttctctct gggtcgggccc cctttcctct ttgaaatact taaacttaat gggaaatcct 720
 taccagacac tgggggtaac atcgcttttt cccaatctca caaatttaca aaccctcagg 780
 ataggaaaatg tagagacttt cagttagata aggagaatag attttgcctgg gctgacttct 840
 ctcaatgaac ttgaaattaa ggcattaagt ctccggaatt atcagtccca aagtctaaaag 900
 tcgatccggc acatccatca cctgactctt cacttaagcg agtctgcttt cctgctggag 960
 atttttgcag atattctgag ttctgtgaga tatttagaac taagagatac taacttggcc 1020
 aggttccagt tttcaccact gcccgtagat gaagtcagct caccgatgaa gaagctggca 1080
 ttccgaggct cggttctcac tgatgaaagc ttttaacgagc tcctgaagct gttgctgtac 1140
 atcttggaac tgtcggagggt agagttcgac gactgtacct tcaatgggct cggcgatttc 1200
 aaccctcgg agtcagacgt agtgagcagc ctgggtaaa tagaaacagt cactatccgg 1260
 aggttgcata tccccagtt ctatttgttt tatgacctga gtactgtcta ttccctctg 1320
 gagaagggtga agcgaatcac agtagagaac agcaaggctt tcctggttcc ctgctcgttc 1380
 tcccagcatt taaaatcatt agaattctta gacctcagcg aaaatctgat ggttgaagaa 1440
 tatttgaaga actcagcctg taagggagcc tggccttctc taaaaacctt agttttgagc 1500
 cagaatcatt tgagatcaat gcaaaaaaca ggagagattt tgctgactct gaaaaacctg 1560
 acctcccttg acatcagcag gaacactttt catccgatgc ccgacagctg tcagtgggca 1620
 gaaaagatgc gcttctgaa tttgtccagt acagggatcc ggggtggtaaa aacgtgcatt 1680
 cctcagacgc tggaggtggt ggatgttagt aacaacaatc ttgactcatt ttctttgttc 1740
 ttgcctcggc tgcaagagct ctatatttcc agaaataagc tgaaaacact cccagatgct 1800
 tcgttgttcc ctgtgttgct ggtcatgaaa atcagagaga atgcagtaag tactttctct 1860
 aaagaccaac ttggttcttt tcccaaactg gagactctgg aagcaggcga caaccacttt 1920
 gtttgcctct gcgaactcct atcctttact atggagacgc cagctctggc tcaaatcctg 1980
 gttgactggc cagacagcta cctgtgtgac tctccgctc gcctgcacgg ccacaggctt 2040
 caggatgccc ggcctccgt cttggaatgt caccaggctg cactgggtgc tggagtctgc 2100
 tgtgcccttc tcctgttgat cttgctcgtg ggtgccctgt gccaccattt ccacggactg 2160
 tggtagctga gaatgatgtg ggcgtggctc caggccaaga ggaagccaa gaaagctccc 2220
 tgcagggacg tttgctatga tgcctttgtt tcctacagtg agcaggattc ccattgggtg 2280
 gagaacctca tgggtccagca gctggagaac tctgaccgc cctttaagct gtgtctccac 2340

aagcgggact tcgttccggg caaatggatc attgacaaca tcatcgattc catcgaagag 2400
 agccacaaaa ctgtgttcgt gctttctgag aacttcgtac ggagcgagtg gtgcaagtac 2460
 gaactggact tctcccactt caggctcttt gacgagaaca acgacgcggc catccttgg 2520
 ttgctggagc ccattgagag gaaagccatt cccagcgcct tctgcaaact gcgcaagata 2580
 atgaacacca agacctacct ggagtggccc ttggatgaag gccagcagga agtgttttgg 2640
 gtaaactctga gaactgcaat aaagtccctag gttctccacc cagttcctga ctccctaac 2700
 taaggctctt gtgacacaaa ctgtaacaaa gtttataagt aacatagaat tgtattattg 2760
 aggatattaa ctatgggttt tgtcttgaat actgttatat aaatatgtga catcaggaaa 2820
 aaaaaaaaaa aaaaaaaaaa 2838

<210> 8
 <211> 200
 <212> DNA
 <213> Mus musculus

<220>
 <223> Targeting vector

<400> 8
 tctctacaaa ccttagtttt gagccagaat catttgagat caatgcaaaa aacaggagag 60
 attttgctga ctctgaaaaa cctgacctcc cttgacatca gcaggaacac tttccatccg 120
 atgcccgaca gctgtcagtg gccagaaaag atgcgcttcc tgaatttgtc cagtacaggg 180
 atccgggtgg taaaaacgtg 200

<210> 9
 <211> 200
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Targeting vector

<400> 9
 agtctgctgt gcccttctcc tgttgatctt gctcgtaggt gccctgtgcc accatttcca 60
 cggactgtgg tacctgagaa tgatgtgggc gtggctccag gccaaagagga agcccagaa 120
 agctccctgc agggacgttt gctatgatgc ctttgtttcc tacagtgagc aggattccca 180
 ttgggtggag aacctcatgg 200

<210> 10
 <211> 613
 <212> DNA
 <213> Mus musculus

<400> 10
 cctgtcctga gtgaccaaca ggtcgtggag ctgattcctg ggggcacggg catcgtggtg 60
 gaatatgagg atcgtctccg tttcatccag ctgggtcgga aggccggcta gaggagagca 120
 aggagcaggt gccagccatg caggcaggtc tgctgaaggt ggtaccacag gctgtgctgg 180
 acttgctgac ctggcaagag ttggaaaaga aggtgtgtgg ggaccagagg cactgtggat 240
 gctctgcgca acgtcactcg gtttgaggac ttcgaacat ctgacacacg cgtgcagtac 300
 ttttgggagg ctctgaacaa tttaccaat gaggaccgga gccgtttcct gcgcttctg 360
 accggccgca gcgcctccct gctcggatct acatttacc agacaaactg ggctatgaga 420
 ccacagacgc cctgcctgag tcttctacct gctccagcac gctcttctc ccacactatg 480
 ccagtgccaa ggtatgcgag gagaagctcc gctacgccgc gtacaactgt gtgagctttg 540
 ttcttcagaa ggcttggcat ggacatccgc tcaatacca ggctttacca gcagccaaac 600
 tactgtcttc tgg 613

<210> 11
 <211> 200
 <212> DNA
 <213> Artificial Sequence

<220>

<223> Targeting vector

<400> 11

```

caatatccac cctgatgctg gcttccatga ctgcacccca ttctgctgg tcagtgagta 60
cagagatgcc ccagtgtacg ctgaaagtag tgtccaactt tcccaggcag gcttcttcct 120
cctccctagg tggcagccat gcaggcaggt ctgctgaagg tggtagcaca ggctgtgctg 180
gacttgctga cctggcaaga                                     200

```

<210> 12

<211> 200

<212> DNA

<213> Artificial Sequence

<220>

<223> Targeting vector

<400> 12

```

ttaccaaatg aggaccggag cegtttctg cgctttgtca cgggccgag cgcctcct 60
gctcggatct acatttacc agacaaactg gggtagtat gactacagag aggggcagaa 120
ggtgcccccc acacacat gcacgcatc gcgtcactcc acaactata atcctttgct 180
cttcccacct tgccgctatg                                     200

```

<210> 13

<211> 626

<212> DNA

<213> Mus musculus

<400> 13

```

agcagcttca attggccaca ctctgcctct tcaggaggcc tggaggata aataataaaa 60
caaaccaggc aggccggtt ggcctgtgg atggcgctg gaccactga taatcccaa 120
ggacaatgac tcgcatgctg gacctgggc tgtttctggc aggcctctc actgtgaaag 180
gtctcctgca ggacagagat gctccagata tgtatgattc tccagtcagg gtccaagagt 240
ggaggggcaa gaaggatgcc cggcagcttg ctgcacaaa catggaattt ggcttcaagc 300
tactgcagag gctggccagc aatagccctc agggaaacat cttcctatcc cactgagca 360
tctccactgc cttctccatg ctgtctctcg gggcccagaa ctccacactg gaggaaatcc 420
gggaaggctt caacttcaaa gagatgtcaa actgggacgt ccatgcggtt ttccactacc 480
ttctccaaa gctcaaccag gagacggagg acacaaagat gaacttaggc aatgctttat 540
ttatggatca gaactgaggc ccaacagagg ttcttgaact ggctaaaatg tgtatgatgc 600
gacatggtgc cattaattcc agactt                                     626

```

<210> 14

<211> 200

<212> DNA

<213> Artificial Sequence

<220>

<223> Targeting vector

<400> 14

```

ggagatgaca aagattcaat gccatcaaga ctctcaaagc actcggctgt ggtgttcagc 60
accctacagc cgtgtcttc tccatgggt ctctcgttg gctaagatgg gagggagcca 120
tggggtcctc cccttcttg ttaacctca gatctgctgc tcaacaggat aatccccaag 180
gacaatgact cgcattgctg                                     200

```

<210> 15

<211> 200

<212> DNA

<213> Artificial Sequence

<220>

<223> Targeting vector

<400> 15

```

ctgccttctc catgctgtct ctcgggggccc agaactccac actggaggaa atccggggagg 60
gcttcaactt caaagagatg tcaaacaggg acgtccatgc ggctttccac taccttctcc 120
acaagctcaa ccaggagacg gaggacacaa agatgaactt aggcaatgct ttatttatgg 180
atcagaagct gaggccccaa                                     200
    
```

<210> 16

<211> 2531

<212> DNA

<213> Mus musculus

<400> 16

```

cgacagcaac ggcgcccccg gaggatgcgg tggagtttgt gctttgctgc atccgtcact 60
gaagaacaaa ataagagtaa aggagacaag ctgcagagca tgggaggctg tggggtcttc 120
tgaaaccttt gctgggcttt ccgcggagca tgagctttta aaacgaattc ttttcaaaga 180
aaccatttgg tgtagctgga aaaatgatac acatgctaaa tgcagcagcc tatcgggtga 240
aatggaccag atccgggtgct gctaaaaggg ctgcctgcct ggtggctgcg gcataatgctc 300
tgaaaaccct ctatcccatc attggcaagc gtttaaagca gcctggccac aggaaggcaa 360
aagcagaagc ttactcgctt gcagagaaca gagaaatact gcattgcacg gagatcatct 420
gtaaaaaacc tgcgcgggga ctaaattgcag cttttttcaa acagctacta gaacttcgga 480
aaatcctctt tccaaaactt gtgaccactg aaacgggggtg gctctgcctc cactcgggtg 540
ctctaattct aagaacattt ctctctatth atgtggctgg tctggatggg aaaatcgtga 600
aaagcatcgt ggaaaagaag cctcggactt tcatcatcaa attaatcaag tggcttatga 660
ttgctatccc tgctaccttt gtcaacagtg ctatcaggta cctggaatgc aaactggcat 720
tggcctttag aactcgctta gttagaccat cctatgagac ctatttcgca aatcagactt 780
attataaggt gataaatatg gatgggagc tgacaaaccc tgaccagtct cttaccgaag 840
acattatgat gttctcgcga tctgtggctc acctgtattc caaccttacc aaacctatth 900
tagatgtcat tctaacctcc tatactctca tccggacagc tacatccaga ggagcaagcc 960
ctatagggcc caccctgtta gcaggacttg tctgtgatgc cactgctaaa gtactgaaag 1020
cttgctcgcc caaatttggg tgcgtgggtg ctgaagaagc ccacaggaaa ggctactctg 1080
ggtatgtcca ctcccgaatc atagccaatg tagaagaaat tgccttctac agaggacata 1140
aggtagaatg gaagcagctg cagaaatgth acaaggctth agcttaccag atgaacctga 1200
ttttatccaa acgtttatgg tacatcatga tagaacaatt ctgatgaag tatgtgtgga 1260
gcagctgtgg actaattatg gtggctatac ccattatcac tgcaacgggc tttgcagatg 1320
gtgatctgga ggatgggtcca aagcaggcta tggttagcga tccggacagag gccttcacca 1380
ctgcccggaa ctacttgccc tctggagctg atgcaattga aaggattatg tcttcataca 1440
aagagatcac tgaactagca ggttatactg ctgagagtata caatatgttc tgggtcttcg 1500
atgaagtgaag gagaggcatt tataagagaa ctgtcactca ggaacctgaa aacctatgca 1560
agcgtggagg taacctggaa ctaccctca gcgacaccct ggccatcaaa ggaacagtha 1620
ttgatgtgga tcatggaatc atttgtgaaa atgttcccat aattacacca gcgggcgaag 1680
tggtggcttc caggctaaac ttcaaagtg aagaagggat gcatctcttg ataactggtc 1740
ccaacggttg tgggaaaagc tctctcttca gaatcttaag cgggctgtgg cctgtgtatg 1800
aaggagtctt ttataaaccg cctccccaac atatgttcta tattccacag aggccatata 1860
tgtctcttgg aagtctccgg gatcaagtca tttaccctga ctacagggat gacatgcgtg 1920
agaaaggtta cactgaccaa gacctagaac gcatcctgca cagcgtgcac ctctaccaca 1980
tagttcaaag agaaggagga tgggatgcag tcatggactg gaaagatgtc ctttccggag 2040
gggagaagca gagaatgggc atggcgcgga tgttttacc aaccgaag tatgcattgc 2100
tggatgaatg taccagtgc gtgagcatcg acgttgaag aaagatattt caggctgcta 2160
ttggggctgg gatttcctca ctctccataa cacacaggcc ttctctgtgg aaataccaca 2220
ctcatctatt acaattcgat ggcgaaggag gctggcgctt tgaacagttg gacactgcta 2280
tccgtttaac gttgagtgag gaaaagcaaa agttggagtc gcagctcgct ggaattccca 2340
aaatgcaaca gagactcaac gaactatgca aaattctggg ggaagactcg gtgctgaaaa 2400
caatccaaac tccagaaaag acatcctaath ttatcttgac atgttttcag ttaccttcta 2460
ggtgaagcct cagagactct ctctttactg catgcagtat gttaagctaa gtgcagagaa 2520
agcaagccgg c                                     2531
    
```

<210> 17

<211> 200

<212> DNA

<213> Artificial Sequence

<220>

<223> Targeting vector

<400> 17

```
cggcatatgc tctgaaaacc ctctatccca tcattggcaa gcgtttaaag cagcctggcc 60
acaggaaggc aaaagcagaa gcttactcgc ctgcagagaa cagagaaata ctgcattgca 120
cggagatcat ctgtaaaaaa cctgcgccgg gactaaatgc agcttttttc aaacagctac 180
tagaacttcg gaaaatcctc                                     200
```

<210> 18

<211> 200

<212> DNA

<213> Artificial Sequence

<220>

<223> Targeting vector

<400> 18

```
tccactcggg ggctctaata tcaagaacat ttctctctat ttatgtggct ggtctggatg 60
ggaaaatcgt gaaaagcatt gtggaaaaga agcctcggac tttcatcatt aaattaatca 120
agtggcttat gattgctatc cctgctacct ttgtcaacag tgctatcagg tacctggaat 180
gcaaactggc attggccttt                                     200
```